DTIC FILE COPY

4

AD-A208 182

DISTRIBUTED INFERENCING FOR CLASSIFICATION

(N00014-87-K-2029)

Final Report 8/1/87 - 9/30/88

Prepared For:

DTIC
ELECTE
MAY 25 1989
S D
D

CENTER FOR APPLIED RESEARCH IN AI

Dr. Lashon B. Booker
Dept. of the Navy, Code 7510
Naval Research Lab
Washington, D.C. 20375-50

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Principal Investigator:

Judea Pearl
UCLA Computer Science Department
Cognitive Systems Laboratory
Los Angeles, CA. 90024-1596
(213) 825-3243

1. SUMMARY OF PAST RESEARCH

In the first year of funding our research focused on classification tasks under uncertainty, with special emphasis on propagating beliefs in a system containing a mixture of categorical and probabilistic relationships. Previously available techniques were able to handle taxonomic trees, where the main relationship is ISA -- i.e., class membership. The knowledge involved in object recognition contains non-binary relationships (e.g., IN-BETWEEN) arranged in non-decomposable structures. The difficulties encountered stem from the incompleteness of the model. In other words, we normally have information about the relationship between a variable and each of its neighbors but not between a variable and all of its neighbors taken together. This precludes the construction of a complete Bayesian model. To overcome this difficulty, we have developed a new formulation of the Dempster-Shafer theory in terms of a static constraint-networks (representing stable knowledge) bombarded by randomly fluctuating constraints, (representing uncertain evidence) (Appendix I). We have also devised a scheme for computing Dempster-Shafer belief functions in that model, using Assumption-Based Truth Maintenance Systems (ATMS) and Incidence Calculus (Appendix I).

As an alternative to the Dempster-Shafer theory, we have investigated another scheme of handling partial models, based on the observation that facts and rules are usually communicated with virtual certainty, without attaching to them numeric measures of belief. Qualitative representations involve simplicity of elicitation, communication, encoding, and computation. They suffer however from the ills of classical monotonic logics which preclude the representation of context-dependent information. In Appendix II we describe a system which retains the context-dependent nature of probabilistic inferences, yet involves non-numeric propositions. Input knowledge is cast in statements with four levels of certainty: true, false, likely and unlikely. However, unlike classical multivalued logics, the system retains the essential properties of probabilistic inferences, such as conditionalization, independence and causality. This allow us to manipulate evidence by logical means while maintaining the probabilistic semantics (hence the plausibility) of the output statements, cast in the same 4-value vocabulary.

Appendix II contains examples of property inheritance relationships, and shows how the inference scheme handles them satisfactorily. The scheme is ripe for applications involving object recognition, but further study is required to identify what relationships are inherited across spatial relationships. For example, only some of the properties of an object are inherited by its parts and, vice versa, only few of the part's properties are inherited by the object as a whole.

We are currently studying another approach to handle the mixture of probabilistic and categorical relationships for object recognition tasks, based on the maximum-entropy principle. This approach attempts to complete the model and, unlike conventional methods of maximum entropy, exploits the "almost categorical" nature of spatial constraints to extract a computationally tractable calculus of property inheritance.

☑
□
□
per ltr
Availability Codes
Dist | Avail and/or Special
A-1
QUALITY INSPECTED 2

2. LIST OF RESULTS

- An axiomatic system (called Graphoids) was developed, which provides a formal characterization of informational dependencies and their graphical representations.
- Techniques were developed for learning causal structures from emprirical data.
- Relevance-based control strategies were developed to optimize the selection of tests and to minimize the network activity during updating.
- The interrelationships between the Bayesian and Dempster-Shafer approaches to uncertainty were given formal definition in terms of provability conditions and constraint networks.
- Probabilistic semantics was developed for a subset of default reasoning, leading to consistency criteria, sound inferences and the qualitative management of causality.
- Sound and complete graphical procedures for identifying the set of parameters needed for answering a given probabilistic query.
- A proof that *d*-separation is the most complete graphical criterion for detecting conditional independencies in influence diagrams.

These results are described in the following publications.

3. RELATED PUBLICATIONS AND REPORTS

Dalkey, N., "A Logic of Information Systems," UCLA Cognitive Systems Laboratory, *Technical Report 870057 (R-98),* August 1987. *Proceedings,* 6th Workshop on Maximum Entropy and Bayesian Networks, Seattle, Washington, August 1987.

Dechter, Rina & Pearl, J., "Network-Based Heuristics for Constraint-Satisfaction Problems," *Artificial Intelligence,* Vol. 34:1 December 1987, pp. 1-38. In L. Kanal and V. Kumar (eds), *Search in AI,* Springer-Verlag, 1988, pp. 370-425.

Pearl, J., "On Logic and Probability," *Computational Intelligence,* Vol. 4, April 1988, pp. 90-94.

Dechter, Rina, & Pearl, J., "Tree-Clustering Schemes for Constraint-Processing," *Proceedings,* AAAI-88, St. Paul, Minnesota, August 1988, pp. 150-154. Also in *Artificial Intelligence,* Vol. 38:3, April 1989, pp. 353-366.

Pearl, J. "Probabilistic Semantics for Inheritance Hierarchies with Exceptions," UCLA Cognitive Systems Laboratory, *Technical Report 870052 (R-93),* July 1987.

Geffner, H., & Pearl, J., "A Framework for Reasoning with Defaults," UCLA Cognitive Systems Laboratory, *Technical Report 870058 (R-94),* March 1988. To appear in *Proceed-*

ings, Society for Exact Philosophy Conference, Rochester, New York, June 1988.

Pearl, J., ''Deciding Consistency in Inheritance Networks,'' UCLA Cognitive Systems Laboratory, *Technical Report 870053 (R-96)*, August 1987.

Pearl, J., ''A Probabilistic Treatment of the Yale Shooting Problem,'' UCLA Cognitive Systems Laboratory, *Technical Report 870068 (R-100)*, September 1987.

Verma, T., & J. Pearl, ''Influence Diagrams and d-Separation,'' UCLA Cognitive Systems Laboratory, *Technical Report 880052 (R-101)*, March 1988.

Geiger, D., ''Towards the Formalization of Informational Dependencies,'' UCLA Cognitive Systems Laboratory, *Technical Report 880053 (R-102)*, (Based on the author's MS thesis), Dec 3, 1987.

Verma, T., ''Some Mathematical Properties of Dependency Models,'' UCLA Cognitive Systems Laboratory, *Technical Report (R-103)*, August 1987.

Pearl, J., ''On Probability Intervals,'' UCLA Cognitive Systems Laboratory, *Technical Report 880094 (R-105)*. January 1988. *International Journal of Approximate Reasoning*, Vol. 2:3, July 1988.

Pearl, J., ''Bayesian and Belief-Functions Formalisms for Evidential Reasoning: A Conceptual Analysis,'' UCLA Cognitive Systems Laboratory, *Technical Report 880054 (R-106-S)*, January 1988. *Proceedings*, 5th Israeli Symposium on Artificial Intelligence, Tel Aviv, December 1988, pp. 398-424.

Pearl, J., ''Evidential Reasoning under Uncertainty,'' UCLA Cognitive Systems Laboratory, *Technical Report 880055 (R-107)*, February 1988. In *Exploring Artificial Intelligence: Survey Talks from the National Conferences on Artificial Intelligence*, Ed. H. Shrobe, Morgan & Kaufmann, 1988, pp. 381-418. To appear in *Annual Review of Computer Science*, Vol. 4, 1989.

Geffner, H., ''On the Logic of Defaults,'' UCLA Cognitive Systems Laboratory, *Technical Report 880058 (R-110)*, March 1988. *Proceedings*, AAAI-88, St. Paul, Minnesota, August 1988, pp. 449-454.

Geiger, D., & J. Pearl, ''On the Logic of Causal Models,'' UCLA Cognitive Systems Laboratory, *Technical Report 880060 (R-112)*, March 1988. *Proceedings*, 4th Workshop on Uncertainty in Artificial Intelligence. Minneapolis, Minnesota, August 1988, pp. 136-147.

Pearl, J., D. Geiger & T. Verma, ''The Logic of Influence Diagrams,'' UCLA Cognitive Systems Laboratory, *Technical Report 880061 (R-114)*, April 1988. To appear in R.M. Oliver and J.Q. Smith (Eds), *Influence Diagrams, Belief Nets and Decision Analysis*, Sussex, England: John Wiley & Sons, Ltd., 1989. A shorter version, (R-114-S), in *Kybernetica*, Vol. 25:2, 1989, pp. 33-44.

Geiger, D. A. Paz & J. Pearl, ''Axioms and Algorithms for Inferences Involving Probabilistic Independence,'' UCLA Cognitive Systems Laboratory, *Technical Report (R-119),* December 1988. Submitted to *Information and Computation,* 1989.

Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* San Mateo: Morgan Kaufmann Publishers, 1988.

4. APPENDICES

I. Pearl, J., ''Bayesian and Belief-Functions Formalisms for Evidential Reasoning: A Conceptual Analysis,'' UCLA Cognitive Systems Laboratory, *Technical Report 880054 (R-106-S),* January 1988. *Proceedings,* 5th Israeli Symposium on Artificial Intelligence, Tel Aviv, December 1988, pp. 398-424.

II. Geffner, H., & Pearl, J., ''A Framework for Reasoning with Defaults,'' UCLA Cognitive Systems Laboratory, *Technical Report 870058 (R-94),* March 1988. To appear in *Proceedings,* Society for Exact Philosophy Conference, Rochester, New York, June 1988.

APPENDIX I.

TECHNICAL REPORT
R-106S-I
November 1988

BAYESIAN AND BELIEF-FUNCTIONS FORMALISMS FOR EVIDENTIAL REASONING: A CONCEPTUAL ANALYSIS (1)(2)

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles

ABSTRACT

This paper clarifies the relationships between two popular formalisms used in evidential reasoning tasks: Bayesian inference and Belief-Function theory (also known as the Dempster-Shafer theory).

Bayesian methods requires the specification of a complete probabilistic model that relates the set of hypotheses to the set of anticipated observations. When a full specification is not available, approximate strategies are devised to complete the model from the information available. Subsequently, the model can be used to answer any probabilistic query, for example, finding the probability of a set of hypotheses, given the evidence, or, finding the most likely set of hypotheses, given the evidence.

In contrast, the Dempster-Shafer (D-S) theory sidesteps the missing specifications and compromises on its answers; it computes probabilities of *necessity*, or *provability*, instead of probabilities of truths. Domain knowledge is represented by categorical constraints of compatibility among the propositions involved, and partial evidence is modeled as randomly fluctuating constraints. These two sets of constraints are then used for assembling proofs of the logical necessity of the conclusions. The stronger the evidence, the more likely it is that a complete proof will be assembled.

The paper illustrates and contrasts these two formalisms using simple examples. It is shown that, while the D-S theory enjoys the advantage of admitting partially specified models, it lacks the flexibility of utilizing useful probabilistic knowledge (when such is available) and may inherit many of the problems associated with monotonic logic.

(1) Sections of this paper are based on [Pearl, 1988a; Chapter 9].

(2) This work was partially supported by the National Science Foundation, Grant #IRI-8610155 and Naval Research Laboratory, Grant #N00014-87-K-2029.

In Proceedings, The 5th Israeli Symposium on AI, Vision and Pattern Recognition, December 1988, pp. 398-424.

I. INTRODUCTION

Evidential reasoning is the process of drawing plausible conclusions from uncertain clues and incomplete information. Such processes permeate almost every field in AI, ranging from diagnosis and forecasting to image interpretation, speech recognition and language understanding. By and large, these tasks have been handled by ad-hoc heuristic techniques, embedded in domain-specific procedures and data structures. Recently, there have been a strong movement to seek a more formal and principled basis for evidential reasoning, and the two most popular contenders that have emerged are the Bayesian and the Dempster-Shafer (D-S) approaches. In machine vision, for example, both the Bayesian (e.g., Binford et al, 1988) and the D-S [e.g., Andress and Kak, 1988] approaches have been actively pursued for various object recognition tasks.

The Bayesian model is by far the more familiar between the two, resting on a rich tradition of probability theory, and statistical decision theory and is supported by excellent axiomatic and behavior arguments. The three defining attributes of the Bayesian approach are (1) reliance on complete probabilistic model of the domain (2) willingness to accept subjective judgments as an expedient substitute for empirical data and (3) the use of Bayes conditionalization as the primary mechanism for updating beliefs in light of new information.

Belief functions offer an alternative to Bayesian inference, in that they do not require the specification of a complete probabilistic model and, consequently, they do not (and cannot) use conditionalization to represent the impact of new evidence. Originally, belief functions were introduced by Dempster [1967] as a generalization of Bayesian inference wherein probabilities are assigned to sets rather than to individual points. In Dempster's formulation, belief functions are interpreted as lower and upper probabilities induced by a family of probability distributions. This interpretation of belief functions, still a favorite of many researchers [Kyburg, 1987], leads to a great deal of confusion as to the empirical basis of the inputs and the semantics of the probability intervals computed.

Shafer [1976] has reinterpreted Dempster's theory as a model of evidential reasoning including two interacting frames: a probabilistic frame representing the evidence, and a frame of possibilities over which categorical compatibility relations are defined. Shafer's reinterpretation abandons the idea that belief functions arise as lower bounds of some family of ordinary probability distributions; rather, it views belief functions as the fundamental components in reasoning about uncertain evidence. The mathematical relations between belief functions, possibility theory, and inner measures are discussed in Shafer [1987], Ruspini [1987] and Fagin and Halpern [1988].

The main purpose of this paper is to offer a new, easily comprehensible interpretation of belief functions, which exposes their underlying semantics and thus facilitates a better understanding of their power and range of applicabili'y vis a vis those of Bayesian inference. Our interpretation of belief functions reflects a transla ion of Shafer's formulation into a language more familiar to an AI audience, appealing to the notions of proofs and constraints. The key element in our interpretation is the theme that belief f inctions represent the probability of *provability*.

namely, the probability that the constraints imposed by the available evidence, together with the constraints which normally govern the environment, will be sufficient for compelling the truth of a proposition and excluding its negation. We shall first demonstrate this interpretation on a simple example, made as crisp as possible, then address the more general issues of computational, epistemological and semantic adequacies.

2. BASIC CONCEPTS

We contrast the Bayesian and the Dempster-Shafer (D-S) theories using the classical Three Prisoners puzzle [Gardner, 1961]. Three prisoners, A, B, and C, await their verdict, knowing that one of them will be declared guilty and the other two released. Prisoner A asks the jailer, who knows the verdict, *to pass a letter to another prisoner—to one who is to be released*; later, the jailer tells Prisoner A that he gave the letter to prisoner B. The problem is to assess Prisoner A's chances of being declared guilty. The problem can be formulated in terms of three exhaustive and mutually exclusive propositions, G_A, G_B, and G_C, where G_i stands for "Prisoner i will be declared guilty." We also have the jailer's testimony, which could have been either "B" or "C," and thus can be treated as a bi-valued variable L (connoting *letter recipient*) taking on the values $\{L_B, L_C\}$.

In the Bayesian treatment of the problem we make two assumptions. First, we assume that lack of prior knowledge regarding the verdict translates to equal prior probabilities on the components of G, namely, $\pi(G_A) = \pi(G_B) = \pi(G_C) = 1/3$. Second, we assume that if G_A were true, the jailer would choose a letter recipient at random, giving equal probabilities to B and C. These two assumptions yield the Bayesian network of Figure 1*a*, where Figure 1*b* depicts the conditional probability matrix $P(L|G)$ necessary for full characterization of the information source L. This model yields the answer $P(G_A \mid L_B) = 1/3$, meaning that the jailer's testimony is totally irrelevant to A's prospects of being released. If, on the other hand, the jailer does not deliver the letter at random but prefers B (or C) with probability $P(L_B \mid G_A) = q$, then $P(G_A \mid L_B)$ will be given by the formula $\frac{q}{1+q}$, which varies smoothly from 0 (if B is avoided) to 1/2 (if C is avoided).

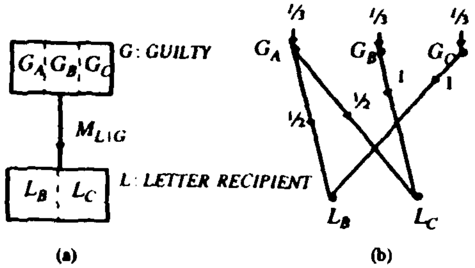


Figure 1. (a) A Bayesian network representation of the Three Prisoners puzzle and (b) the conditional probability matrix characterizing the link $G \rightarrow L$.

These two assumptions are avoided in the D-S theory. Having no idea what sort of trial the prisoners were given, we know only that *any one of the prisoners can be the guilty one*, i.e., none can be ruled out conclusively. Similarly, not knowing the process by which the letter recipient was chosen, all we can assert with certitude is that L_B is *compatible* with both G_A and G_C and is incompatible with G_B (assuming the jailer is honest). Thus, following the testimony L_B, the only possible states of affairs are the two combinations $\{(G_A, L_B), (G_C, L_B)\}$; all others are ruled out. These legal states are called *extensions* in the language of logic, *solutions* in the language of constraint processing [Montanari 1974], *tuples* in the language of relational databases, *possibilities* in fuzzy logic [Zadeh 1981], and *frame of discernment* in the Dempster-Shafer theory. The constraints that determine which extensions are legal are called *compatibility relations* (e.g., that exactly one prisoner will be found guilty), representing items of information that one chooses to cast in hard, categorical terms because a more refined model is unavailable (e.g., regarding the unlikely circumstances under which all three prisoners will be pardoned).

Clearly, not having the parameters $\pi(G_i)$ and $P(L_j \mid G_i)$ keeps us from constructing a complete probabilistic model of the story and keeps us from answering probabilistic queries of the type "How certain is G_A in light of the jailer's testimony?"—previously encoded as $P(G_A \mid L_B)$. In the partial model available, the probability $P(G_A \mid L_B)$ could be anywhere from 0 to 1, depending on the prior probability π. On the other hand, if by *certainty* we mean the assurance that G_A can be *proved* true, then the certainty of G_A, logically speaking, is zero.

The Dempster-Shafer theory stands between these two extremes, claiming that even in the logical interpretation of certainty, the assurance that there exists a proof for a proposition A can take on various strengths, depending on the strength of the evidence available (namely, how close it is to inducing a logical proof of A). This degree of assurance is called a *belief function* and is denoted by $Bel(A)$.† In our story, both $Bel(G_A)$ and $Bel(\neg G_A)$ are zero, because having total ignorance regarding the trial and verdict process means we have no evidence capable of enabling a logical proof of either G_A or $\neg G_A$.

Under what conditions will these belief functions be anything but zero? One obvious condition is when the negation of a proposition becomes incompatible with the evidence. For example, since G_B is incompatible with L_B, we have $Bel(\neg G_B) = 1$, stating that $\neg G_B$ is compelled by the evidence. But the more interesting condition occurs when *partial* evidence becomes available in favor of some propositions. For example, if the jailer says, "Gee, I forgot who got the letter—I think it was B, but I am only 80 percent sure," we can no longer prove $\neg G_B$. Yet, taking the jailer's testimony literally, we could say that there is an 80% chance that his memory is correct, compelling the truth of $\neg G_B$, so $Bel(\neg G_B) = 0.80$. Similarly, if we have good reason to believe that the witnesses in the trial gave equal support to each prisoner's guilt and that the verdict reflects this testimony fairly, then (and only then) we can take the liberty of assigning equal *weights* to the components of G.

† $Bel(A)$ is to be distinguished from $BEL(A)$, which, in the Bayesian literature, is defined by $BEL(A) \triangleq P(A \mid \text{all evidence})$.

Let us first focus on the case of equal weights, ignoring for the moment the jailer's information. The weight distribution process can be modeled as a chance event: Imagine a switch that oscillates randomly among three positions, G_A, G_B, and G_C, and in each of these positions assigns the value *TRUE* to the corresponding proposition and no value to the others (Figure 2*a*). The other propositions in the system are affected by the switch only indirectly, via compatibility relations they must satisfy with G_A, G_B, and G_C. These relations are shown schematically by the graph in Figure 2*a*. For example, the link connecting G_B and $\neg G_C$ indicates that the two may coexist, while the absence of a link between G_A and G_B means those two are incompatible. (Not all of the compatible pairs are shown in the graph; for example, $\neg G_A$ and $\neg G_B$ are compatible whenever G_C is true.)

If we are asked now about the chance of G_A being provable, the answer will be ⅓, because the switch spends one-third of the time in position G_A, where the truth of G_A is established externally. During this time G_B and G_C can be proved false by virtue of being incompatible with G_A. 'hus, averaging over all three positions of the switch,

$$Bel(G_A) = Bel(G_B) = Bel(G_C) = 1/3$$

and

$$Bel(\neg G_A) = Bel(\neg G_B) = Bel(\neg G_C) = 2/3,$$

exactly as in the Bayesian treatment.

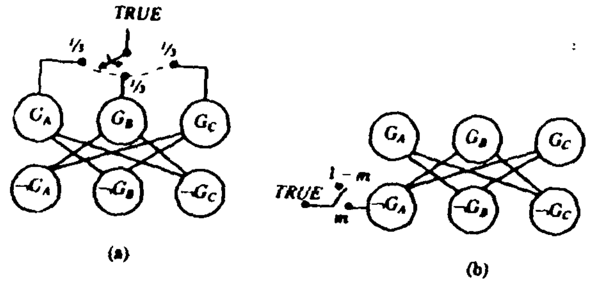


Figure 2. Random switch model representing (a) equal prior probabilities, and (b) an alibi weakly supporting *A*'s innocence.

The departure from the Bayesian formalism comes when we devise fancier mechanisms for the weight-distribution switch so as to form more faithful models of how (according to D-S philosophy) people reason with partial evidence. Assume, for example, that the evidence gathered during the trial is not available to us in its entirety, but rather we have access to a small portion of it, namely, an alibi weakly supporting Prisoner *A*'s innocence. Assume, further, that the alibi bears exclusively on *A*'s whereabouts at the time of the crime and has no direct bearing on *B*'s or *C*'s involvement. The D-S theory will model this case by the switch shown in Figure 2*b* : a fraction *m* of the time, the switch will force the truth of $\neg G_A$, while the remaining $1 - m$ of the time it will stay in a neutral position, lending support to no specific hypothesis or, equivalently, supporting the universal hypothesis $\theta = G_A \vee G_B \vee G_C$.

To calculate the belief functions $Bel(G_A)$ and $Bel(\neg G_A)$ we first identify the positions of the switch in which G_A can be proved true, then calculate the percentage of time spent in these positions. In the first position, representing a convincing alibi, the switch forces the truth of $\neg G_A$, while in the neutral position it is compatible with both G_A and $\neg G_A$ so nothing can be proved. Hence, $Bel(\neg G_A) = m$ and $Bel(G_A) = 0$. The belief acquired by the other elementary propositions is zero (prior to the jailer's testimony) because even in the first position the switch is compatible with each of the four propositions: G_B, G_C, $\neg G_B$ and $\neg G_C$.

The parameter $m(A)$, measuring the strength of the argument in favor of a proposition *A*, is called the *basic probability assignment*, and the proposition *A* upon which an argument bears directly is called the *focal element*. If there is only one focal element *A*, then the weight $1 - m(A)$ is assigned to the universal proposition θ, and the belief in any other proposition *B* is given by

$$Bel(B) = \begin{cases} 1 & \text{if} \quad B = \theta \\ m(A) & \text{if} \quad A \supset B \\ 0 & \textit{otherwise}. \end{cases} \qquad (1)$$

A complex piece of evidence may be represented by a switch with more than two positions, each position forcing a different constraint on the knowledge base for a certain fraction of the time. For example, if evidence was found suggesting that the guilty person was either left-handed (with weight m_1) or black-haired (with weight m_2) but not both, and if Prisoners *A* and *B* are left-handed while *B* and *C* have black hair, the constraint $G_A \vee G_B$ will be imposed a fraction m_1 of the time, $G_B \vee G_C$ will be imposed a fraction m_2 of the time, and the rest of the time, $1 - m_1 - m_2$, no external constraint will be imposed.

In general, if there are several focal elements *A*, the total weight still sums to unity,

$$\sum_A m(A) = 1, \qquad (2)$$

and $Bel(B)$ may be affected by all the A's, via

$$Bel(B) = \sum_{A,\, A \supset B} m(A). \qquad (3)$$

The summation reflects the fact that if B can be proved from several *mutually exclusive* assumptions (represented by certain positions of the switch) then $Bel(B)$, the probability that B is provable, is the total weight assigned to those assumptions (corresponding to the time that the switch spent in those positions).

The measure $1 - Bel(\neg A)$ is called the *plausibility* of A, denoted

$$Pl(A) = 1 - Bel(\neg A), \qquad (4)$$

representing the probability that A is compatible with the available evidence, i.e., that it cannot be disproved and is therefore *possible*. In our example, $Pl(G_A) = 1 - m$, while G_B and G_C have plausibility 1. The interval

$$Pl(A) - Bel(A) = 1 - [Bel(A) + Bel(\neg A)] \geq 0 \qquad (5)$$

represents the probability (fraction of time) that both A and $\neg A$ are compatible with the available evidence.

3. COMPARING BAYESIAN AND DEMPSTER-SHAFER FORMALISMS

We see that the D-S theory differs from Bayes' theory in several aspects. First, it accepts an incomplete probabilistic model when some parameters (*e.g.*, the prior or conditional probabilities) are missing. Second, the probabilistic information that is available, like the strength of evidence, is interpreted not as likelihood ratios but rather as random epiphenomena that impose truth values to various propositions for a certain fraction of the time. This model permits a proposition and its negation simultaneously to be compatible with the switch for a certain portion of the time, and this may permit the sum of their beliefs to be smaller than unity. Finally, given the incompleteness of the model, the D-S theory does not pretend to provide full answers to probabilistic queries but rather resigns itself to providing partial answers. It estimates how close the evidence is to forcing the truth of the hypothesis, instead of estimating how close the hypothesis is to being tru..

This last point is the most important departure between the two formalisms and is best illustrated, in the Three Prisoners puzzle, by trying to incorporate the jailer's information L_B into the equal-weight model $m(G_A) = m(G_B) = m(G_C) = ⅓$ (see Figure 3). Starting with $Bel(G_A) = ⅓$, we now ask for the revised value of $Bel(G_A)$ given L_B, i.e., the proportion of the time that proposition G_A is provable, considering all of the available evidence. Clearly, the time spent by the switch in position G_B is incompatible with the evidence L_B, so we exclude this time from the calculation. The remaining ⅔ of the time is divided equally between G_A and G_C, hence G_A is forced to be true with probability ½, yielding $Bel(G_A) = Bel(\neg G_A) = ½$. The Bayesian analysis gave $P(G_A \mid L_B) = ⅓$, assuming a random choice model ($P(L_B \mid G_A) = ½$, as in Figure 1) and $P(G_A \mid L_B) = \frac{q}{1+q}$ for a partial model with uncertain $q = P(L_B \mid G_A)$.

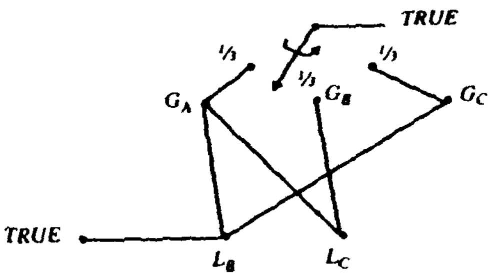


Figure 3. The D-S representation of the Three Prisoners story, incorporating equal prior probabilities and the evidence $L_B = TRUE$.

This disparity is not surprising in view of the fact that we still have an incomplete probabilistic model on our hands, as the process by which B was selected remains unspecified. It is quite possible that the jailer's choice was not random but was marked by a deliberate attempt to avoid choosing C whenever possible. Under such extreme circumstances, the jailer's answer L_B could be avoided only ⅓ of the time (when B is guilty), indeed leaving A and C with equal chances of being the condemned prisoner. What may sound somewhat counterintuitive is that from all the scenarios that could be used to complete the model, D-S theory appears to select a rather extreme and unlikely one, which also happens to be the one that puzzle books warn us to avoid [Gardner 1961].†

Actually, the D-S theory never attempts to complete the model, and if it appears to be doing so, it is only as an occasional by-product of the way weights are distributed. The disparity between the answers produced by the two formalisms stems from the semantics of the answers. While the Bayesian approach interprets "belief in A" to mean the conditional probability that A is true, given the evidence e, the D-S approach calculates the probability that the proposition A is *provable* given the evidence e and given that e is consistent. Due to the inconsistency of L_B with the evidence we previously had in favor of G_B, e is consistent only ⅔ of the time, and within this time, G_A can be proved with probability ½. Thus, instead of the conditional probability $P(A \mid e)$, the D-S theory computes the probability of the logical entailment $e \models A$. The en-

† An even stranger result obtains in the Thousand Prisoner version of the story [Pearl 1988a, Chapter 2] where D-S theory gives ½ and the Bayesian model 1/1000.

tailment $e \models A$ is not a proposition in the ordinary sense, but a meta-relationship between e and A, requiring a logical, object-level theory by which a proof from e to A can be constructed. In the D-S scheme the object-level theory consists of categorical *compatibility* relations among the propositions, stating, for example, that L_B is compatible with G_A but incompatible with G_B.

Remarkably, whereas calculating the probability $P(A \mid e)$ (as well as the probability of the material conditional, $P(e \supset A)$) requires a complete probabilistic model, calculating $P(e \models A)$ does not. For example, in the incomplete model of Figure 3, $P(L_B \models G_A)$ can be calculated as ⅓ without any assumption about the process by which the letter recipient was selected; we simply take 1 minus the (normalized) weight assigned to all propositions compatible with both L_B and the negation of G_A—namely, 1 minus the (normalized) time the switch spends at G_C.

At this point, it is natural to ask whether conditional probability information, if available, can be incorporated in the D-S model, and whether it will lead to the same answer as the Bayesian model. The answer is a qualified yes; it will, provided that the information is sufficient for forming a complete probabilistic model. Instead of dealing with individual propositions, we now create the set of all feasible extensions(*) and attach to each extension a weight m equal to the appropriate joint probability dictated by the probabilistic model. To illustrate, if in the Three Prisoners example we accept the equal-prior random-selection model, then the four feasible extensions $\{(G_A, L_B), (G_A, L_C), (G_B, L_C), (G_C, L_B)\}$ initially receive the weights $\{1/6, 1/6, 1/3, 1/3\}$ (see Figure 1*b*). This assignment can be modeled by a four-position switch whose contacts represent extensions rather than atomic propositions. When the evidence $e = L_B$ is obtained, it rules out two extensions, (G_A, L_C) and (G_B, L_C), and forces the switch to spend ⅓ of its time at (G_C, L_B) and 1/6 of the time at (G_A, L_B). Thus,

$$Bel(G_A) = \frac{1/6}{1/3 + 1/6} = 1/3, \quad Bel(\neg G_A) = \frac{1/3}{1/3 + 1/6} = 2/3,$$

as in the Bayesian analysis.

We see from this example that any complete probabilistic model can be encoded in the D-S formalism, albeit in a somewhat clumsy manner. Probabilities are encoded as weights assigned to individual extensions, instead of conditional probabilities among propositions. This might not seem a severe limitation when we are processing complete models, but it hinders the handling of partial models. Large fragments of empirical knowledge cast in the form of conditional probabilities (such as the relation between symptoms and diseases) cannot be incorporated into the D-S compatibility frame until we have sufficient information to form a complete probability model and to calculate the weights of individual extensions. In the Three Prisoners story, for example, even if we obtain ample evidence that the jailer acts randomly (i.e., $P(L_B \mid G_A) = 1/2$), we cannot incorporate this evidence so as to affect $Bel(G_A)$ as long as we are missing the prior probability $P(G_A)$. This is because conditional probabilities cannot be modeled as switches that constrain the set of possible extensions. In other words, the statement $P(A \mid B) = p$ cannot be converted into an equivalent statement of the form $P[f(A, B)] = q$, where f is some Boolean function of A and B [Goodman 1987].

To a certain degree this limitation also applies to Bayesian methods; we can begin drawing inferences only when the model is complete. However, in cases where we have acquired a large body of knowledge and are missing just a few parameters, the Bayesian approach encourages us to assume *any* reasonable values for the missing parameters, so that the knowledge acquired will not be totally wasted. For example, assuming equal priors in the Three Prisoners story enables us to deploy the information about the jailer's behavior $P(L_B \mid G_A) = q$, and conclude $P(G_A \mid L_B) = q / (1 + q)$. The D-S analysis continues to conclude $Bel(G_A) = 0$, regardless of the jailer's behavior, as long as the prior $\pi(G_A)$ remains unspecified.

Thus, while the D-S approach has the capability to tolerate total ignorance, it lacks flexibility to utilizing partial information when such is available.

TRUTH, POSSIBILITY, AND NECESSITY

Compatibility constraints represent a simplified black-and-white abstraction of a world that is ridden with exceptions. The notion of provability normally reflects mathematical artifice rather than empirical reality. Thus, the preceding discussion might leave the erroneous impression that the D-S theory deals with a totally unrealistic model of the world and provides answers to contrived, uninteresting queries about such a model. There are, however, cases where compatibility relations are natural representations of world knowledge, and where queries concerned with the probability of provability rather than the probability of truth are the ones that we wish to ascertain.

For example, suppose we face the problem of scheduling classes; we have a set of teachers, a set of topics, a set of classrooms, and a set of time slots, and our task is to cover all of the topics with the available resources. Before we actually select an assignment (if one exists), our knowledge is represented in the form of constraints: some teachers cannot teach certain topics, some classrooms are unavailable at certain times, etc. On the basis of this knowledge, all we can answer are questions about possibilities, e.g., whether it is possible to take topic x in time slot y, whether it is feasible to get teacher z for topic w, etc. To answer such queries, we need to search the space of feasible assignments (i.e., those satisfying all of the constraints) and test whether there is an assignment that satisfies the query. If a query Q is satisfied by at least one feasible assignment, we say that Q is *possible*; if it is satisfied by all feasible assignments, we say it is *necessary* or *provable*.

Probabilistic measures enter in the form of partial constraints. For example, suppose one of the teachers states, "Due to a medical problem, there is an 80 percent chance that I will be unable to teach Mondays and Thursdays." This information imposes on the problem an additional constraint—this time probabilistic in nature, similar to the action of the switch in Figure 3. It assigns a weight of $m = 0.80$ to a narrower set of feasible assignments and a weight of $m = 0.20$ to

(*) We use the term "extensions" to denote what probabilists call "elementary events" or "points" and what logicians call "models" or "worlds".

the original set, in which the teacher was presumed to be available on Mondays and Thursdays. It is natural now for some other teacher to ask, "What are the chances that I will be able to teach mathematics?" or, "What are the chances that I will be forced to teach on Tuesday evening?" These queries, concerning possibilities and necessities, translate directly into the D-S measures of $Pl(Q)$ and $Bel(Q)$, respectively. To answer such queries, we must solve the assignment problem for each state of the switch, determine for each state whether the query is provable, possible, or impossible, and then compute the desired probability using the weight m.

Can such answers be computed by a pure, single-level Bayesian model? Not really. We could of course assume a uniform distribution over the feasible assignments in each state, combine the two distributions with the appropriate weights—0.20 and 0.80—and then calculate the resultant probability of the query sentence Q (e.g., that teacher x will be assigned to teach on Tuesdays). However, this is not the same as the probability that Q is possible, or the probability that it is necessary, and the difference might be significant. For example, the inquirer might have a strong aversion to teaching on Tuesdays and hence be determined to talk his way out of any such assignment if a feasible alternative exists. He is, therefore, concerned not with the probability of being assigned a Tuesday time slot, but rather with the probability that such an assignment will be necessary for lack of an alternative. In such cases, the purely probabilistic approach will not provide the desired answer. Queries regarding probability of possibility (or necessity) require two levels of knowledge and hence cannot be answered by treating the compatibility constraints as probability statements having value 0 or 1. Rather, the compatibility constraints must remain outside the probabilistic model. They serve as object-level theories which, in themselves, are assigned probability measures, thus rendering the necessity of a query a random event.

4. DEMPSTER'S RULE OF COMBINATION

When several pieces of evidence are available, their impacts are combined by assuming that the corresponding switches act independently of each other. For example, if in addition to Prisoner A's alibi (Figure 2b) the trial records include testimony supporting A's guilt to a degree m_2, one can imagine two random switches operating simultaneously and asynchronously, the first as described in Figure 2b and the second spending a fraction m_2 of the time constraining G_A to the value *TRUE* and staying neutral the rest of the time (see Figure 4a).

Clearly, for a fraction m_1m_2 of the time the two switches are in conflict with each other— one is constraining G_A to *TRUE*, and the other is constraining $\neg G_A$ to *TRUE*— thus permitting no consistent extension. For a fraction $(1-m_1)m_2$ of the time G_A is true while switch 2 is neutral, rendering $G_A, \neg G_B$, and $\neg G_C$ provable. Similarly, for a fraction $m_1(1-m_2)$ of the time $\neg G_A$ is true while switch 1 is neutral, rendering $\neg G_A$ (but no other proposition) provable. Summing up and normalizing by the time of no conflict, $1-m_1m_2$, we have

$$Bel(G_A) = Bel(\neg G_B) = Bel(\neg G_C) = \frac{m_2(1-m_1)}{1-m_1m_2},$$

$$Bel(\neg G_A) = \frac{m_1(1-m_2)}{1-m_1m_2},$$

$$Bel(G_B) = Bel(G_C) = 0$$

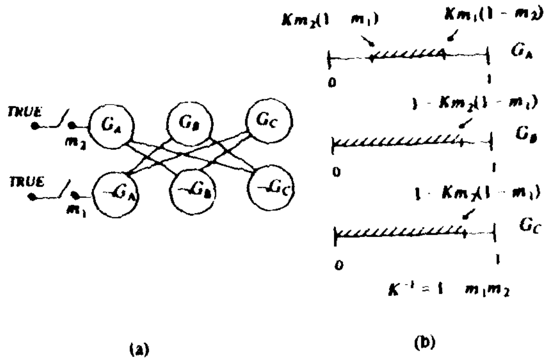


Figure 4. Combination of two sources of evidence in the Three Prisoners puzzle. (a) m_1 and m_2 represent the percentage of time each switch is closed. (b) Belief intervals for the three propositions G_A, G_B, and G_C.

These measures are shown schematically in Figure 4b. The assumption of evidence independence, coupled with the normalization rule above, leads to an evidence-pooling procedure known as *Dempster's rule of combination*. The combined impact of several pieces of evidence can be calculated, again, by computing the fraction of time that a given proposition A is compelled to be true by the combined action of all switches, assuming that they operate independently. Thus, the analysis of belief functions amounts to analyzing the set of extensions permitted by a network of static constraints (representing generic knowledge), subject to an additional set of externally imposed, fluctuating constraints (representing the impact of the available evidence). For any combination of the evidential constraints, we need to examine the set of extensions permitted by that combination and decide whether the proposition A is entailed by the set; i.e., whether every extension contains A and none contains $\neg A$. The total time that a system spends under constraint combinations that compel A, divided by the total time spent in no-conflict combinations, yields $Bel(A)$.

The constraint-network formulation of Dempster's combination rule is illustrated schematically in Figure 5. It shows a static network of variables X, Y, Z, V ... (the nodes) interacting via local constraints (the arcs), subject to the influence of two switches that impose additional, fluctuating constraints on various regions of the network. To illustrate the analysis of the extension sets, let us assume that the static network represents the classical graph-coloring problem: Each node may take on one of three possible colors, 1, 2, or 3, but no two adjacent nodes may take on identical colors. The positions of the switches represent additional constraints, e.g., C_{XY} means either X or Y must contain the color 1, while C_Z means Z cannot be assigned the color 2. The relative time that a switch spends enforcing each of the constraints is indicated by the weight measures $m_1(C_X)$, $m_1(C_{XY})$, $m_2(C_Z)$, etc. Our objective is to compute $Bel(A)$ and $Pl(A)$, where A stands for the proposition $V = 1$, namely, variable V is assigned the color 1.

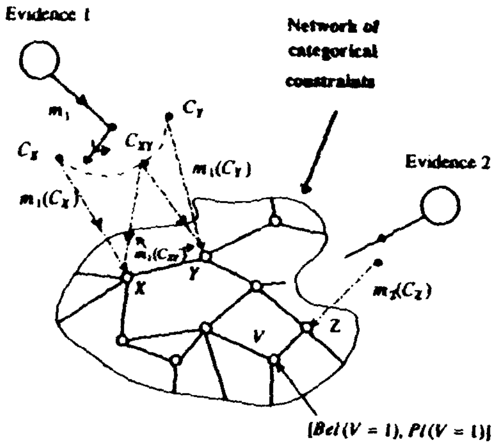


Figure 5. Multiple evidence modeled as random switches imposing additional constraints on a static network of compatibility relations.

Figure 6 represents typical sets of solutions to the coloring problem under different combinations of the switches (the values shown are fictitious). Each row represents one extension (or solution), where the entries indicate the values assigned to the variables (columns). The first set of solutions is characterized by having the value 1 assigned to V in every row. If the system spends a fraction α of the time in such combinations of switches, we say that $P[e \models V = 1] = \alpha$, namely, the proposition $A: V = 1$ can be proved true with probability α, given the evidence e. A type-2 position is characterized by the column of V containing 1's as well as alternative values, e.g., 2 and 3. Each such position (or position combination) is compatible with both A and $\neg A$. Similarly, a type-3 position permits only extensions that exclude $V = 1$, while a type-4 position represents a conflict situation: There exists no extension consistent with all the constraints. $Bel(A)$ and $Pl(A)$ are computed from the time spent in each type of constraint combination:

$$Bel(A) = \frac{\alpha}{\alpha + \beta + \gamma}\ , \tag{6}$$

$$Pl(A) = 1 - Bel(V \neq 1) = 1 - \frac{\gamma}{\alpha + \beta + \gamma} = \frac{\alpha + \beta}{\alpha + \beta + \gamma}\ . \tag{7}$$

These are illustrated as a belief interval in Figure 6b.

The preceding analysis can be rather complex. The graph-coloring problem, even with only three colors, is known to be NP-complete. Moreover, if each piece of evidence is modeled by a two-position switch and we have n such switches, then a brute force analysis of $Bel(A)$ will require solving 2^n graph-coloring problems. Listing the solutions obtained under all switch combinations and identifying the combinations that yield $e \models A$ seems hopeless. Fortunately, these difficulties can sometimes be alleviated by exploiting certain topological properties of the constraint network. The latter is done by decomposing the network into a tree (a join tree), where solutions can be obtained in linear time [Dechter and Pearl 1987, 1988]. Adaptations of

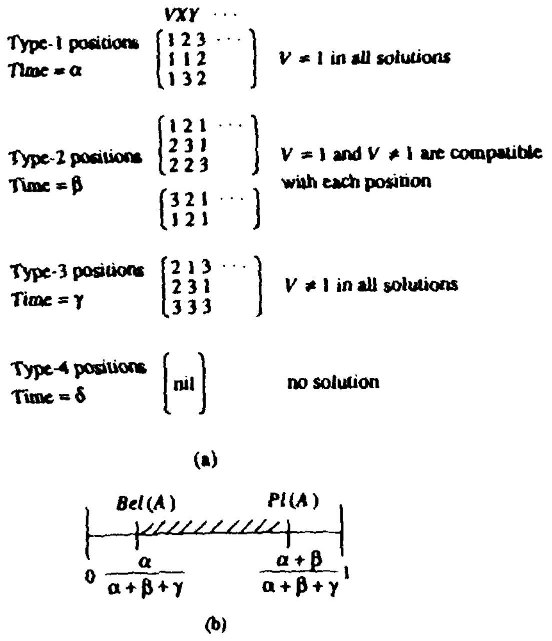


Figure 6. a) Four types of constraints in the graph coloring problem and b) the resulting belief interval for the proposition $A: V = 1$.

tree decomposition to belief function computations are reported in Kong [1986] and Shafer, Shenoy, and Mellouli [1987].

5. THE MEANING OF BELIEF INTERVALS

At this point, it is worthwhile to reflect on the significance of the interval $Pl(A) - Bel(A)$ in the D-S formalism. This interval is often *interpreted as the degree of ignorance we have about probabilities*, namely, the range were the "true" probability should fall if we had a complete probabilistic model. *Such measures of ignorance might be a useful supplement to Bayesian methods, which always provide point probabilities and thus can give a false sense of security in the model.*

Unfortunately, the D-S intervals have little to do with ignorance; nor do they represent bounds on the probabilities that might ensue once ignorance is removed. This was already demonstrated in the Three Prisoners puzzle. We saw that despite our total ignorance regarding the process by which the jailer chose the letter recipient, the interval $Pl(G_A) - Bel(G_A)$ was zero, thus *giving the false impression that the answer* $Bel(G_A) = ½$ was based on a complete model (with the jailer avoiding C whenever possible). *At the same time*, knowledge of the selection process should sway the posterior probability $P(G_A \mid e)$ all the way from 0 to ½.

The disappearance of the "ignorance" interval is not an isolated incident but will occur whenever a piece of evidence imparts all of its weight to singleton hypotheses. *In other words*, if e' induces $Pl'(A) = Bel'(A)$, then regardless of the ignorance we possessed before and regardless of any ignorance that might be conveyed by future evidence, $Pl(A) - Bel(A)$ will remain zero forever. In particular, whenever we start with a complete probabilistic model (where all belief intervals are zero), no amount of conflicting evidence will ever succeed in widening that interval so that it reflects the conflict.

The upshot is that many sources of *ignorance or uncertainty* about probabilities are not represented in the D-S formalism. In particular, the uncertainty caused by high sensitivity to unknown contingencies [Pearl, 1988a; Chapter 7] cannot be represented by belief intervals. For example, suppose we know that a given coin was produced by a defective machine— precisely 49% of its output consists of double-headed coins, 49% are no-headed coins, and the rest are fair. This description constitutes a complete probabilistic model which predicts that the outcome of the next toss will be heads with probability 50% and warns us that the prediction is extremely susceptible to new information regarding the coin. Though most people would hesitate to commit a point estimate of 50% to the next outcome of the coin, the D-S theory nevertheless assigns it a belief of 50%, with zero belief interval. Now imagine that we toss the coin twice and observe tails and then heads. This immediately implies that the coin is fair, and most people would regain confidence that the next toss has a 50% chance of turning up heads. Yet this narrowing of the confidence interval remains unnoticed in the D-S formalism; the theory will again assign the next outcome a belief of 50%, with zero belief interval.

The disappearance of the difference $Pl - Bel$ in the Three Prisoners puzzle is a byproduct of the normalization used in Dempster's rule. Avoiding this normalization would have yielded an interval [⅓, ⅔] for G_A, reflecting the fact that G_A and $\neg G_A$ can each be proved only one-third of the time (assuming no proposition is truly provable from a contradiction). Indeed, normalization by the time-without-conflict stands at odds with the basic definition of Bel as the probability of necessity. Instead of this intended probability, the normalized version of Bel reflects the *conditional* probability of *necessity given that the set of extensions is nonempty*. Valuable information is sometimes lost in this conditionalization process. After all, *a state of contradiction represents an inadequacy* in our model of the world (e.g., that the jailer is reliable, or that only two prisoners will be released), not a major flaw in the world itself. So the support that a proposition receives in conflicting situations depends on *how we plan to extend or refine* the model once a contradiction is found. A more reasonable approach would be to keep two intervals, the one measuring the degree of conflict and the other measuring the degree of noncommitment. That would *entail characterizing* each proposition by four parameters corresponding to the four types of solution sets (see Figure 6). (Similar criticism of the normalization used in the D-S approach was advanced by Zadeh [1984].)

6. APPLICATIONS TO RULE-BASED SYSTEMS

Representing knowledge in the Dempster-Shafer theory is more natural *when the compatibility* relations are expressed in rule form. A rule, r, is a constraint among a group of propositions, expressed in *IF-THEN* format:

$$r:\ a_1 \wedge a_2 \wedge \cdots \wedge a_n \Rightarrow c\ . \tag{8}$$

Propositions $a_1, \ldots, a_n$ are called the *antecedents* (or *justifications*) of the rule, and c is its *consequent*. The semantics of the rule lies in forbidding *any extension in which the antecedents are all true while the consequent* is false; in other words, a rule is equivalent to the constraint

$$r:\ \neg[a_1 \wedge a_2 \wedge \cdots \wedge a_n \wedge \neg c]\ . \tag{9}$$

Normally, rules are based on tacit *assumptions*, the failure of which (called *exceptions)* may invalidate the rule. For example, I may assert the rule r: "If it is Sunday John will go to the baseball game," tacitly assuming the prerequisites "John still holds season tickets," "John is not sick," etc. Since *such* assumptions are too numerous to explicate, they are often summarized by giving the rule a measure of strength, m. For example, the rule above might be given a strength of $m = 0.80$, *indicating* 80% certainty that none of the implicit exceptions will materialize.

One of the attractive features of the D-S formalism is that it allows rules with exceptions *to be treated* much the same as rules of inference in deductive logic. A rule is treated as just another *compatibility constraint in the knowledge base*, while m measures the strength with which the constraint is enforced.† Using our random switch metaphor, the strength m translates to a switch that spends a fraction m of the time imposing the constraint conveyed by the rule. The activity of the switch during the time remaining depends on the nature of the exceptions anticipated. Some exceptions (e.g., "John is sick") lead to the negation of the conclusion while others (e.g., "John has no season tickets this year") render the conclusion unknown or uncommitted. In the first case the switch will force the negation of the conclusion while in the second case it will spend its remaining time in a neutral position. Thus, the rule author must be aware of the type of assumptions summarized by the rule strength. For the sake of simplicity we will characterize rules with a simple switch model that supports the consequence c but not its negation. More sophisticated models, such as three-position switches, yield similar results.

COMBINING BELIEF FUNCTIONS IN RULE NETWORKS

Assume we have a system of rules and a list F of observed facts called *premises*, and we wish to find the belief $Bel(\cdot)$ attributable to some proposition c. This amounts to computing the probability that a proof exists from the premises in F to the conclusion c. Each proof consists of a sequence of rules $r_1, r_2, ..., r_m$ such that the antecedents in each r_i are either premises or are proved and the consequence of r_m is the desired conclusion c. Graphically, a proof can be represented by a directed graph like the one shown in Figure 7*a*, where the root nodes are all premises, the leaf node is c, and each bundle of converging arrows represents a given rule. The arcs connecting the arrows represent the logical AND function between the antecedents of each rule.

The collection R of all rules available to a system can be represented by an AND/OR graph like the one in Figure 7*b*, where an OR function is understood to exist between any two parent bundles converging toward the same node. The graph in Figure 7*b* contains two proofs for c, (r_1, r_2) and (r_3, r_4). If t can no longer be asserted as a premise, the proof (r_1, r_2) is no longer valid, but c can still be proved via (r_3, r_4).

We are now in a position to calculate $Bel(c)$, namely, the probability that c is provable in a system of uncertain rules, where each rule r_i is characterized by a strength measure m_i. A system of such rules is equivalent to an AND/OR graph whose links are interrupted by random switches, as shown in Figure 8. The task of computing $Bel(c)$, then, amounts to calculating the percentage of time that some proof graph of c remain uninterrupted. In the special case where every rule has a single antecedent, the problem can be reduced to finding the percentage of time that an uninterrupted path exists between some premise and the conclusion. Such problems have been studied extensively in the area of network reliability, and in general they turn out to be NP-hard, even under the assumption that the interruptions are independent of each other [Rosenthal, 1975].

† Since this treatment of rules is the prevailing view among D-S practitioners, we shall call it the *D-S approach*. An alternative treatment, viewing rules as conditional probability statements, is in principle also permitted within the D-S formalism, but because it requires a complete probabilistic model we shall call it the *Bayesian approach*.

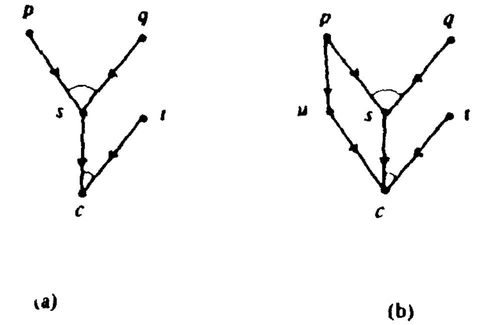


Figure 7. (a) A proof graph for proposition c, representing two rules $p \wedge q \rightarrow s$, $s \wedge t \rightarrow c$ and the premises p, q, t. (b) An AND/OR graph representing four rules.

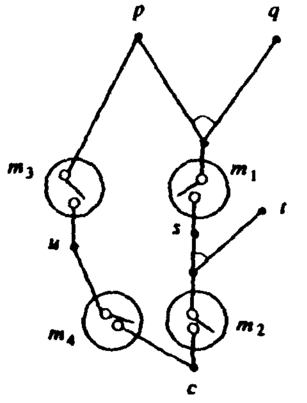


Figure 8. Random switch model for the rule network of Figure 8*b*.

A brute-force way of calculating $Bel(c)$ would be to enumerate all switch combinations, test each combination to see if a proof exists for it, and then total the time the switches spend in combinations that pass the test. For a system with n rules, this would require the enumeration of 2^n combinations. Fortunately, the simple nature of the network in Figure 8 permits the calculation to be done without enumerating all combinations. Since the network contains two disjoint proofs, the active times of the two proof graphs are mutually independent; hence, the time that c

is unprovable is equal to the product of the amounts of time that each proof graph is inactive, i.e., $(1 - m_3 m_4)(1 - m_1 m_2)$. The rest of the time c is provable, hence

$$Bel(c) = 1 - (1 - m_3 m_4)(1 - m_1 m_2). \tag{10}$$

Note that instead of enumerating all $2^4 = 16$ switch positions, we had to enumerate only the two proof paths (in general, proof graphs) (r_1, r_2) and (r_3, r_4), calculate the active time t_i of each path, and then calculate $Bel(c)$ using the formula

$$Bel(c) = 1 - \prod_i (1 - t_i). \tag{11}$$

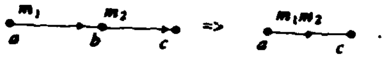
Such shortcuts are not feasible in general rule networks. For example, if we add the rule $r_5{:}s \rightarrow u(m_5)$ to the system of Figure 7b, an additional proof graph is added, $\{r_1, r_5, r_4\}$, whose activation time is dependent on the other proof graphs, and we can no longer calculate $Bel(c)$ by multiplying the inactive times of each separate proof graph. Rather, we must enumerate all of the distinct ways that one or more proof graphs remain active, i.e.,

$$Bel(c) = [1 - (1 - m_3 m_4)(1 - m_1 m_2)] + m_1 m_5 m_4 (1 - m_2)(1 - m_3). \tag{12}$$

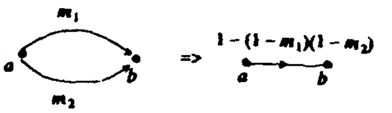
The first term represents the condition that at least one of the proofs, (r_1, r_2), (r_3, r_4), is active, and the second represents the proof remaining under the complementary condition.

What permits shortcuts such as the one taken in Figure 7b is a topological feature called *series-parallel*. This feature allows recursive solution of many graph problems including network flow, network reliability, belief functions, and probabilities. Formally, a rule network is said to be series-parallel if it can be reduced to a single rule by repeated application of the following two operations:

1. Series reduction:



2. Parallel reduction:



It is clear from this definition that series-parallel rule networks permit the calculation of belief functions in time proportional to the number of rules (as opposed to the number of switch combinations), since each reduction operation reduces the number of rules by one. For example, the network of Figure 10 can be reduced in three operations, yielding

$$Bel(c) = m_1(m_2 \| m_3 m_4) = m_1[1 - (1 - m_2)(1 - m_3 m_4)]. \tag{13}$$

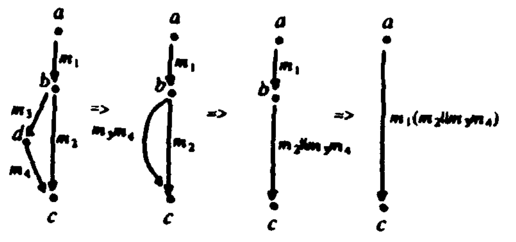


Figure 10. Reducing a series-parallel rule network to a single rule ($m_1 \| m_2$ stands for $1 - (1 - m_1)(1 - m_2)$).

In general networks, however, the calculation of $Bel(c)$ may require exponential time.

THE LIMITS OF SYNTACTIC ANALYSIS

Though it involves only simple graph operations, the foregoing analysis is at variance with most uncertainty management techniques used in rule-based systems. These techniques are based on local syntactic analysis, where the uncertainty associated with the consequent of each rule is presumed to be solely a function of the uncertainties of the antecedents of the rule and the uncertainty of the rule itself. In the D-S formulation of syntactic techniques these uncertainties are represented by an interval $[Bel, Pl]$, so the interval associated with the consequent of each rule is presumed to be a sole function of the intervals that characterizes the rule and the interval that characterize the antecedents [Ginsberg 1984; Falkenhainer 1986; Baldwin 1987]. Moreover, when two rules converge toward the same conclusion, the interval associated with the conclusion is determined by combining the intervals associated with the individual rules. This is precisely where syntactic systems deviate from the principles dictated by the D-S theory. Such combination functions are normally derived under the assumption that the items of evidence conveyed by the rules would be independent of each other. Applying this combination function to every pair of converging arrows in a large network may violate this independence assumption, especially if proof paths overlap. For example, given the truth of a in the initial network of Figure 10, syntactic analysis will compute $Bel(c)$ as follows:

$$Bel(b) = m_1,$$

$$Bel(d) = m_3\, Bel(b) = m_3 m_1,$$

$$Bel(c) = m_4\,Bel(d) \parallel m_2\,Bel(b) = m_4 m_3 m_1 \parallel m_2 m_1$$

$$= 1 - (1 - m_4 m_3 m_1)(1 - m_2 m_1)$$

$$= m_1(m_4 m_3 + m_2 - m_4 m_3 m_2 m_1) \,. \qquad (14)$$

The correct result is

$$Bel(c) = m_1(m_2 \parallel m_3 m_4) = m_1(m_2 + m_3 m_4 - m_2 m_3 m_4),$$

as in Eq. (13). The difference between the two expressions is equal to $m_2 m_3 m_4(m_1 - m_1^2)$, and it clearly stems from counting the arc m_1 twice. An extensional system is too local to realize that the beliefs at b and d originate from the same source.

It is easy to find conceptual examples that amplify the discrepancy between the two approaches and thus highlight the conditions under which extensional systems lead to paradoxical conclusions. To maximize the difference $m_1 - m_1^2$, we let $m_1 = ½$ and $m_2 = m_3 = m_4 = 1$, and assemble the following system of rules:

r_1: If I flip the coin (a), then it will turn up heads (b) ($m_1 = ½$)

r_2: If the coin turns up heads (b), then you win (c) ($m_2 = 1$),

r_3: If the coin turns up heads (b), then I lose (d) ($m_3 = 1$),

r_4: If I lose (d), then you win (c) ($m_4 = 1$).

Suppose I flip the coin ($a = TRUE$); what is the belief attributable to your victory (c)? The correct answer is clearly ½, since the path $b \to d \to c$ is superfluous. Yet the answer computed by an extensional system is $Bel(c) = ¾$, as if my loss contributes an extra piece of evidence toward *your winning*.

In general, rule-based syntactic techniques are easier to program and faster to run, as they require no network supervision and permit each rule to fire independently. The problem is that they violate the independence assumption when two or more proof paths share a common origin; in rule networks that are tree structured, this error condition disappears. Thus, we conclude that D-S analysis, like the Bayesian formulation [Pearl, 1986], admits syntactic techniques only when the rules form a tree structure [Hájek 1987]. Belief functions computations using assumption based truth maintenance systems (ATMS) are discussed in [Pearl 1988a].

Interestingly, a Bayesian analysis will produce the same result as Eqs. (10) and (11) under the following assumptions:

1. A rule $r : a \to b\ (m)$ is interpreted as two conditional probability statements: $P(b \mid a) = m$ and $P(b \mid \neg a) = 0$;

2. Converging rules interact disjunctively, via the noisy-OR model [Pearl, 1988a; Section 4.3.2].

These two assumptions permit the construction of a complete probabilistic model (i.e., a Bayesian network) for any acyclic rule network. The probabilities $BEL(A) = P(A \mid e)$ calculated from such models are identical to the belief functions $Bel(A)$ calculated from the D-S model, for any proposition A *in the rule set*. However, the negations of these propositions get the probabilities $BEL(\neg A) = 1 - BEL(A)$ whereas in the D-S model they are assigned zero $Bel(\cdot)$ values ($\neg A$ cannot be proved by any rule set unless $\neg A$ appears as a consequent of at least one rule).

7. BAYES VS. DEMPSTER-SHAFER: A SEMANTIC CLASH

The essential difference between the Bayesian and D-S interpretations of the rules shows up in systems that have a mixture of conflicting rules, some supporting a proposition A and some supporting its negation, $\neg A$. In such systems the semantic gap between the two approaches leads to qualitatively different conclusions. Whereas the D-S scheme resolves conflicts by Dempster's normalization, the Bayesian approach resolves them by a more cautious mechanism, appealing to their conditional probability interpretation. As a result, the D-S approach will inherit all of the problems of classical monotonic logic when applied to situations requiring belief revision. We shall demonstrate these problems using a simple three-rule example, the so-called penguin triangle.

Consider the rule set R:

r_1: $p \to \neg f$ (m_1), meaning "Penguins normally don't fly;"

r_2: $b \to f$ (m_2), meaning "Birds normally fly;"

r_3: $p \to b$ ($m_3 = 1$), meaning "Penguins are birds."

To emphasize our strong conviction in these rules, we make m_1 and m_2 approach unity and write

$$m_1 = 1 - \varepsilon_1 \,, \quad m_2 = 1 - \varepsilon_2 \,,$$

where ε_1 and ε_2 are small positive quantities. Assume we find an animal called Tweety that is categorically classified as a bird and a penguin, and we wish to assess the likelihood that Tweety can fly. In other words, we are given the premises p and b, and we need to compute $Bel(f)$ using the D-S approach or $P(f \mid p, b)$ using the Bayesian approach.

The Bayesian approach, treating rules as conditional probabilities, immediately yields the expected result—that Tweety's "birdness" does not render her a better flyer than an ordinary penguin. The reason is that the entailment $p \supset b$ permits us to replace $P(f \mid p, b)$ by $P(f \mid p)$, giving

$$P(f \mid p,b) = P(f \mid p) = 1 - P(\neg f \mid p) = 1 - m_1 = \varepsilon_1 . \tag{15}$$

In the D-S approach, on the other hand, if we treat the rules as a system of uncertain compatibility constraints (see Eqs. (8) and (9)), a counterintuitive result emerges: birdness seems to endow Tweety with extra flying power. This is shown in Table 1, where the four states of the rules r_1 and r_2 are enumerated along with the associated probabilities and the provability state of the proposition *Fly*.

Table 1.

Probabilities	r_1	r_2	*Fly*	$\neg$*Fly*
$\varepsilon_1\varepsilon_2$	inactive	inactive	not provable	not provable
$(1-\varepsilon_1)\varepsilon_2$	active	inactive	not provable	provable
$\varepsilon_1(1-\varepsilon_2)$	inactive	active	provable	not provable
$(1-\varepsilon_1)(1-\varepsilon_2)$	active	active	conflict	conflict

Summing over the states where *Fly* is provable, and normalizing, we obtain

$$Bel(Fly) = \frac{\varepsilon_1(1-\varepsilon_2)}{1-(1-\varepsilon_1)(1-\varepsilon_2)} = \frac{\varepsilon_1 - \varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2-\varepsilon_1\varepsilon_2} \approx \frac{\varepsilon_1}{\varepsilon_1+\varepsilon_2} . \tag{16}$$

We see that the belief attributable to Tweety's flying critically depends on whether she is a penguin-bird or just a penguin. In the latter case, rule r_1 dictates $Bel(Fly) = \varepsilon_1$, which is negligibly small. In the former case, adding the superfluous information that all penguins are birds and birds normally fly makes $Bel(Fly)$ substantially higher, as in Eq. (16). It does not go to zero with ε_1 and ε_2, but depends on the relative magnitudes of these quantities. If the proportion of nonflying birds (ε_2) is smaller than the proportion of flying penguins (ε_1), Tweety's flying will be assigned a belief measure greater than 0.50. Using switches with *FALSE-TRUE* positions or with *FALSE-NEUTRAL-TRUE* positions to model rules yields essentially identical results.

Identical results are also obtained when rule r_3 is not asserted with absolute certainty ($m_3 = 1$) but is subject to exceptions, i.e., $m_3 = 1 - \varepsilon_3 < 1$. The Bayesian analysis yields

$$P(f \mid p,b) \le \frac{\varepsilon_1}{1-\varepsilon_3} , \tag{17}$$

meaning that as long as ε_3 remains small, penguin-birds have very little chance of flying, regardless of how many birds cannot fly (ε_2). The D-S analysis, on the other hand, still yields the paradoxical result

$$Bel(f) \cong \frac{\varepsilon_1}{\varepsilon_1+\varepsilon_2} , \tag{18}$$

meaning that if nonflying birds are very rare, i.e., $\varepsilon_2 \approx 0$, then penguin-birds have a very big chance of flying.

The clash with intuition revolves not around the exact numerical value of $Bel(f)$ but rather around the unacceptable phenomenon that rule r_3, stating that penguins are a subclass of birds, plays no role in the analysis. Knowing that Tweety is both a penguin and a bird renders *Bel(Tweety flies)* solely a function of m_1 and m_2, regardless of how penguins and birds are related. This stands contrary to common discourse, where people expect class properties to be overridden by properties of more specific subclasses.

While *in classical logic the three rules in our example would yield an unforgivable con*tradiction, the uncertainties attached to these rules, together with Dempster's normalization, now render them manageable. However, they are managed in the wrong way whenever we interpret if–then rules as randomized logical formulas of the material-implication type, instead of statements of conditional probabilities. The material-implication interpretation of if–then rules is so fundamentally wrong that it cannot be rectified by allowing exceptions in the form of randomization. The real source of the problem is the property of transitivity, $(a \to b, b \to c) \Rightarrow a \to c$, which is basic to the material-implication interpretation. There are occasions when rule transitivity *must be totally suppressed, not merely weakened, to avoid get*ting strange results. One such occasion occurs in property inheritance, where subclass specificity should override superclass properties. Randomization, in this case, weakens the flow of inference through the chain but does not bring it to a dead halt, as it should.

This phenomenon also arises outside the realm of property inheritance. For example, consider these rules:

r_1: If I am sick, then I can't answer the door (m_1),

r_2: If I am home, then I can answer the door (m_2),

r_3: If I am sick, then I stay home ($m_3 = 1$).

Rule r_3 tells us that exceptions to rule r_2, due to sickness, are already reflected in the measure m_2, and exceptions to rule r_1, including those emanating from staying home, have been summarized in the measure m_1. Thus, given that I am sick, the conclusion is that I cannot answer the door, with confidence m_1; given that I am both sick and at home, the same conclusion applies, and the same confidence, too.

In abductive tasks, rule transitivity can lead to even stranger results. Consider the example:

r'_1: If the ground is wet, then it rained last night (m'_1),

r'_2: If the sprinkler was on, then the ground is wet ($m'_2 = 1$).

If we find that the ground is wet, rule r'_1 tells us that $Bel(Rain) = m'_1$. Now, suppose we learn that the sprinkler was on. Instead of decreasing $Bel(Rain)$ by explaining away the wet ground, the new evidence leaves $Bel(Rain)$ the same. More seriously, suppose we first observe the sprinkler. Rule r'_2 will correctly predict that the ground will get wet, and without even inspecting the ground, r'_1 will conclude that it rained last night, with $Bel(Rain) = m'_1 m'_2$.

These difficulties have haunted nonmonotonic logic for years (see [Pearl, 1988b; Ginsberg, 1987] for more details) and will be inherited by the D-S analysis whenever it treats if–then rules as material implications, however much weakened by randomization. The problems can be circumvented by two methods, neither of which is truly satisfactory. One method requires the rule author to state explicitly the exceptions (or assumptions) underlying each rule. For example, rule r'_2 will be phrased "If I am at home, I can answer the door, unless I am sick, or asleep, or under a gun threat ... in which case I will not be able to answer the door." This method works well under the D-S analysis, but the enormous number of potential exceptions to each rule prevents it from being practical. The second method, used in inheritance systems [Touretzky 1986; Etherington 1987], is to use extra logical criteria to decide when transitivity is applicable. For example, Ginsberg has proposed the meta-rule "Never apply a rule to a set when there is a corresponding rule which can be applied to a subset" [Ginsberg 1984]. It can be shown [Pearl, 1988a; Chapter 10] *that such priorities among rules emerge automatically if the rules are* simply given their proper interpretation, namely, conditional probability statements with probabilities close to 1.

CONCLUSIONS

The current popularity of the D-S formalism stems both from its willingness to admit partially specified models and its compatibility with the classical, proof theoretical style of logical inference, sharing the syntax of deductive databases and logic programming. However, inattentive use of this similarity with logic may cause reasoning based on the D-S theory to inherit many of the problems associated with monotonic logic. Among these we find:

1. Inability to infer conditional beliefs, thus compromising decisions which rely on the actual occurence of events, as opposed to their necessity.

2. Inability to accomodate partial probabilistic information in the form of conditional probabilities or assumption-based modes of reasoning.

3. Improper handling of information in the form of default statements.

The Bayesian model, on the other hand, while fairing well on the three counts above, requires the specification of a complete probabilistic model before reasoning can commense. When only a partial specification is available, inferences must rely on approximate model-completion strategies. The computational efforts required by the two formalisms are roughly the same, with a slight advantage to Bayesian methods.

REFERENCES

Andress, K. and Kak, A. 1988. Evidence Accumulation and Flow of Control. *AI Magazine*, Vol. 9:2, Summer 1988, pp. 75-94.

Baldwin, J. F. 1987. Evidential support logic programming. *Fuzzy Sets and Systems* 24:1-26.

Dechter, R., and Pearl, J. 1987. Network-based hueristics for constraint- satisfaction problems. *Artificial Intelligence* 34 (no. 1): 1-38.

Dechter, R., and Pearl, J. 1988. Tree-clustering schemes for constraint-processing. *Proc., 7th Natl. Conf. on AI (AAAI-88)*, Minneapolis, Minnesota, pp. 150-154; to appear in *Artificial Intelligence*, 1989.

Dempster, A. P. 1967. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statistics* 38:325-39.

Etherington, D. W. 1987. More on inheritance hierarchies with exceptions: Default theories and inferential distance. *Proc., 6th Natl. Conf. on AI (AAAI-87)*, Seattle, 352-57.

Fagin, R. and Halpern, J. Y. 1988. Uncertainty, Belief and Probability. *Research Report RJ 6191*, IBM Research, Almaden Research Center, 650 Harry Rd., San Jose.

Falkenhainer, B. 1986. Towards a general purpose belief maintenance system. *Proc., 2nd Workshop on Uncertainty in AI*, Philadelphia, 71-76.

Gardner, M. 1961. *Second Scientific American book of mathematical puzzles and diversions*. New York: Simon and Schuster.

Ginsberg, M. L. 1984. Non-monotonic reasoning using Dempster's rule. *Proc., 3rd Natl. Conf. on AI (AAAI-84)*, Austin, 126-29.

Ginsberg, M. L., ed. 1987. *Readings in non-monotonic reasoning*. Los Altos, Calif: Morgan Kaufmann.

Goodman, I. R. 1987. A measure-free approach to conditioning. *Proc. 3rd Workshop on Uncertainty in AI*, Seattle, 270-277.

Hájek, P. 1987. Logic and plausible inference in expert systems. *Proc. of AI Workshop on Inductive Reasoning*, Roskilde, Denmark.

Kong, A. 1986. Multivariate belief functions and graphical models. Ph.D diss., Department of Statistics, Harvard University.

Kyburg, H. E. 1961. *Probability and the logic of rational belief*. Middleton, Conn.: Wesleyan University Press.

Kyburg, H. E. 1987. Bayesian and non-Bayesian evidential updating. *Artificial Intelligence* 31: 271-94.

Levitt, T. S., Binford, T. O., Ettinger, G. J. and Gelband, P. 1988. Utility-Based Control for Computer Vision. *The 4th Workshop on Uncertainty in AI*, Minneapolis, Minnesota, August 1988, pp. 245-256.

Montanari, U. 1974. Networks of constraints, fundamental properties and applications to picture processing. *Information Science* 7:95-132.

Pearl, J. 1986. Fusion, propagation and structuring in belief networks. *Artificial Intelligence* 29 (no. 3):241-88.

Pearl, J. 1988a. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan-Kaufman, San Mateo, CA.

Pearl, J. 1988b. Embracing causality in formal reasoning. *Artificial Intelligence* 35 (no. 2):259-71.

Rosenthal, A. 1975. A computer scientist looks at reliability computation. In *Reliability and Fault Tree Analysis* (eds. Barlow, Fussel and Singpurwalla), Philadelphia: SIAM, pp. 133-152.

Ruspini, E. 1987. Epistemic logics, probability and the calculus of evidence. *Proc., 10th Intl. Joint Conf. on AI (IJCAI-87)*, Milan, 924-31.

Shafer, G. 1976. *A mathematical theory of evidence*. Princeton: Princeton University Press.

Shafer, G. 1987. Belief functions and possibility measures. In *Analysis of fuzzy information*, ed. J. Bezdek, vol. 1: Mathematics and logic, 51-84. Boca Raton, Fla.: CRC Press.

Shafer, G., Shenoy, P. P., and Mellouli, K. 1987. Propagating belief functions in qualitative Markov trees. Working Paper No. 190, School of Business, University of Kansas (Lawrence). To appear in *Intl. Journal of Approximate Reasoning*.

Touretzky, D. 1986. *The mathematics of inheritance systems*. Los Altos, Calif.: Morgan Kaufmann.

Zadeh, L. A. 1981. Possibility theory and soft data analysis. In *Mathematical frontier of the social and policy sciences*, ed. L. Cobb and R. M. Thrall, 69-129. Boulder, Colo.: Westview.

Zadeh, L. A. 1984. Review of *A mathematical theory of evidence*, by Glen Shafer. *AI Magazine* 5 (no. 3):81.

SCALED PATTERN MATCHING

Amihood Amir[*]
Department of Computer Science
and
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742
Tel. (301)454-4134
email address: amir@mimsy.umd.edu

ABSTRACT

The Exact String Matching with Scaling Problem has as its input, a text string of length n and a pattern string of length m. The output is all occurrences of the pattern in the text, *scaled* to all natural multiples, i.e. all appearances of the pattern in all possible natural *sizes* (twice as large, 3 times as large, ..., $\lfloor n/m \rfloor$ times as large) are to be found. This problem is useful for some tasks in computer vision.

We present a simple, optimal $O(n)$ time algorithm for this problem, which scans the text only once.

We extend the problem to two dimensions, where the text is an $n \times n$ matrix and the pattern is an $m \times m$ matrix. Our algorithm finds all scaled appearances of P in T in time $O(mn^2)$.

[*] This research was partially supported by NSF grant CCR-8803641.

APPENDIX II.

To appear in Proceedings, Society for Exact Philosophy Conference, June 1988, Rochester, NY.

Technical Report
R-94-III
June, 1988
CSD-870058

A Framework for Reasoning with Defaults

Hector Geffner
hector@cs.ucla.edu

Judea Pearl
judea@cs.ucla.edu

Cognitive Systems Lab.
Dept. of Computer Science
UCLA

Abstract

A new system of defeasible inference is presented. The system is made up of a body of six rules which allow proofs to be constructed very much like in natural deduction systems. Five of the rules are shown to possess a sound and clear probabilistic semantics that guarantees the high probability of the conclusion given the high probability of the premises. The sixth rule appeals to a notion of irrelevance; we explain both its motivation and use.[1]

1 Motivation

Belief commitment and belief revision are two distinctive characteristics of common sense reasoning which have so far resisted satisfactory formal accounts. Classical logic for instance, cannot accommodate belief revision: new information can only add new theorems. Probability theory, on the other hand, has difficulties in accommodating belief commitment: propositions are believed only to a certain degree which dynamically changes with the acquisition of new information.

Recent years have witnessed a renovated effort to enhance these formalisms in order to overcome such limitations. Those working within the probabilistic framework have tried to devise 'acceptance rules' to work on top of a body of probabilistic knowledge, as to create a body of believed, though defeasible, set of propositions (e.g. [Levi 80]). Those working

[1] This paper is a revised version of [Geffner *et. al.* 87].

within the logic framework have developed 'non-monotonic' inference systems [AI Journal 80] based on classical logic, in which old 'theorems' can be defeated by new 'axioms'.

In comparison, the probabilistic approach has enjoyed a significant advantage over the logical approach. A body of probabilistic knowledge together with an acceptance rule uniquely determines the conclusions that can be derived. Both the probabilistic knowledge base and the acceptance rule can be modified so as to capture those conclusions that appear reasonable. Non monotonic logics, on the other hand, have lacked such *clear semantics*. Not only it has been difficult to tune the set of defeasible rules so as to 'entail' the desired conclusions [see Hanks and McDermott 86], but it has even been difficult to characterize what the conclusions sanctioned by a body of 'defaults' ought to be (see [Touretzky et al. 87], "A clash of *intuitions* ...").

On the positive side, as noted by [Glymour *et. al.* 84] and [Loui 87a], the logical approach has shown that a *qualitative* account of non-monotonic reasoning, which does not require either 'acceptance rules' or the expense and precision of computing with numbers, might be possible, and has even suggested ways in which such an account can be provided.

In this paper we attempt to show that it is possible to combine the best of both worlds. We present a system of defeasible inference which operates very much like natural deduction systems in logic and, yet, can be justified on probabilistic grounds. The resulting system is closely related to the logic of conditionals developed by Adams [Adams 66], as we interpret defaults of the form $P \rightarrow Q$ as constraining the conditional probability of Q given P to be infinitesimally close to one. On the other hand, the appeal to a notion of relevance in our formulation bears a close relationship to those approaches which investigate defeasible reasoning as resulting from the interaction of competing arguments (e.g. [Touretzky 84; Poole 85; Loui 87b; Pollock 87]).

The structure of the paper is as follows. In section 2 we define the core of the system, discuss the need for providing an account of the notion of irrelevance, and present such an account. In section 3 we illustrate the applicability of the system proposed by going through an standard set of examples. We summarize the main contributions in section 4.

2 A System of Defeasible Inference

2.1 The Core

The logic we shall present will be referred as **L** and will be characterized by a set of rules of inference, in the style of natural deduction systems. The goal of **L** is to sanction as theorems the highly likely consequences that follow from a given context. A context $\Gamma = E_K$ is defined by a background context K, which expresses generic knowledge relevant to the domain of discourse, and an evidential set E, which expresses the particular facts which characterize the particular situation of interest. A background context K, $K = \langle L, D \rangle$, is built from a set of closed wffs L and a set D of defaults, represented by meta-linguistic expressions of the form $p \rightarrow q$, where p and q are closed wffs. The evidential set E is given by a collection of closed wffs. We use default schemas of the form $P(\mathbf{x}) \rightarrow Q(\mathbf{x})$, where P and Q are wffs with free variables among those of $\mathbf{x} = \{x_1, \ldots, x_n\}$, to represent the infinite collection of defaults that results from substituting $\mathbf{x}$ by a vector $\mathbf{a}$ of ground terms.

The system of inference implicitly defines the set of conclusions h that follow from a given context E_K, with $K = \langle L, D \rangle$. We write $E \mid\!\sim_K h$ to denote that sentence h is derivable from context E_K in **L**. Likewise, $E, \{f\} \mid\!\sim_K h$, abbreviated $E, f \mid\!\sim_K h$, states that h is derivable from the context that results from adding the sentence f to E. We shall use the notation $\phi(E)$ to refer the sentence that obtains by conjoining the sentences in E. The symbol $\vdash$ stands for derivability in classical first order logic. The initial set of rules we are going to consider is the following:

Rule 1 (Defaults) If $f \rightarrow h \in D$ then $f \mid\!\sim_K h$

Rule 2 (Logic theorems) If $L \cup E \vdash h$ then $E \mid\!\sim_K h$

Rule 3 (Triangularity) If $E \mid\!\sim_K f$ and $E \mid\!\sim_K h$ then $E, f \mid\!\sim_K h$

Rule 4 (Bayes) If $E \mid\!\sim_K f$ and $E, f \mid\!\sim_K h$ then $E \mid\!\sim_K h$

Rule 5 (Disjunction) If $E, f \mid\!\sim_K h$ and $E, g \mid\!\sim_K h$ then $E, f \vee g \mid\!\sim_K h$

Rule 1 permits us to conclude the consequent of a default when its antecedent is all that has been learned. Rule 2 states that theorems that logically follow from a set of formulas can be concluded in any context containing those formulas. Rule 3 permits the incorporation of an established conclusion to the current evidence set, without affecting the status of any other derived conclusions. Rule 4 says that any conclusion that follows from a context whose evidence set was augmented with a conclusion established in that context, also follows from the original context alone. Finally, rule 5 permits reasoning by cases.

Rules 2–5 can be shown to share the inferential power of the system of rules proposed by Adams in [Adams 66] for deriving what he calls the probabilistic consequences of a given set of conditionals. They also appear, in different form, in the logic of conditionals developed by Stalnaker, Lewis and others during the seventies[2]. Interestingly, in the context of non-monotonic logics, Gabbay [Gabbay 85] has also proposed a minimal set of rules which includes rules 3 and 4 above.

We proceed now to investigate some of the properties of the system of defeasible inference defined by rules 1–5. Later on, we shall discuss some of its limitations as we enhance the system with a sixth rule which attempts to capture the notion of irrelevance.

[2] See [Nute 84] for a review.

2.1.1 Some Meta-Theorems

Theorem 1 (Logical Closure) If $E \mid\!\sim_K h$, $E \mid\!\sim_K h'$, and $E, h, h' \vdash h''$, then $E \mid\!\sim_K h''$.

By rule 3, we obtain $E, h \mid\!\sim_K h'$. From rule 2, we get $E, h, h' \mid\!\sim_K h''$. Applying rule 4 twice, the theorem is proved.

Theorem 2 (Equivalent Contexts) If $E \mid\!\sim_K h$ and $\phi(E) \equiv \phi(E')$ then $E' \mid\!\sim_K h$.

Since $E \vdash \phi(E')$, by applying rules 2 and 3 we get $E, E' \mid\!\sim_K h$; which together with $E' \vdash \phi(E)$ and rules 2 and 4, leads to $E' \mid\!\sim_K h$.

Theorem 3 (Exceptions) If $E \mid\!\sim_K h$ and $E, f \mid\!\sim_K \neg h$ then $E \mid\!\sim_K \neg f$.

From $E, f \mid\!\sim_K \neg h$, we can obtain by theorem 1, $E, f \mid\!\sim_K \neg h \vee \neg f$. On the other hand, from rule 2 we can conclude $E, \neg f \mid\!\sim_K \neg h \vee \neg f$. Combining these two results by means of rule 5 and theorem 2, we get $E \mid\!\sim_K \neg h \vee \neg f$ and, therefore, $E \mid\!\sim_K \neg f$ by virtue of theorem 1 and $E \mid\!\sim_K h$.

Some non-theorems:

$E \vdash f$ and $f \mid\!\sim_K h$ does not necessarily imply $E \mid\!\sim_K h$
$E \mid\!\sim_K h$ and $E' \mid\!\sim_K h$ does not necessarily imply $E, E' \mid\!\sim_K h$

Note that the first non-theorem is clearly undesirable. If accepted, it would endow our system with monotonic characteristics of classical logic, precluding exceptions like non-flying birds, etc. The second one would incorrectly authorize to conclude for instance, that John will be happy when married to both a Jane and Mary, on the grounds that he will be happy when married to either one of them.

As we shall see later, the system of rules 1–5 defines an extremely conservative non-monotonic logic. In fact, the inferences sanctioned by these rules do no involve any type of assumptions regarding information absent from the background context. To illustrate this fact, let $K = \langle L, D \rangle$ and $K' = \langle L', D' \rangle$ denote two background contexts, such that $K \leq K'$, i.e. $L \subseteq L'$ and $D \subseteq D'$. We have the following theorem:

Theorem 4 (K-monotonicity) If $E \mid\!\sim_K h$ and $K \leq K'$ then $E \mid\!\sim_{K'} h$.

This theorem follows easily by induction on the minimal length n of the derivation of $E \mid\!\sim_K h$. If $n = 1$, it means that h was derived from E in K either by rule 1 or by rule 2. In either case it is easy to show that h can be derived from E in K'. Let us assume now that h is derivable from E in K in n steps, $n > 1$, and that the theorem holds for all the proofs with length $m < n$. Clearly the last step in the derivation must involve one of the rules 3–5. In any case, the antecedents of such rule must be derivable in a number of steps smaller than n and, therefore, by the inductive assumption, they are also derivable in K', from which it follows that, using the same rule, h is also derivable from E in K'.

Finally, rules 1–5 can be shown to be probabilistically sound. That is, if we interpret defaults of the form $p \rightarrow q$ as constraining the conditional probability of q given p to be infinitesimally close to 1, we can show that each rule preserves the interpretation which assigns a conditional probability $P(h|E, L)$ infinitesimally closed to 1 to expressions of the form $E \mathrel{|\!\!\!\approx_K} h$. We omit the proof here, and refer the interested reader to [Adams 66] and [Pearl *et. al.* 88].

We now turn our attention to an example that shows how the body of rules introduced so far can account for simple patterns of non-monotonic reasoning.

2.2 Example

Example 1. Let us consider the background context $K = \langle L, D\rangle$ depicted in fig. 1 with $L = \{\forall x. penguin(x) \supset bird(x)\}$, $D = \{penguin(x) \rightarrow \neg flies(x), bird(x) \rightarrow flies(x)\}$.

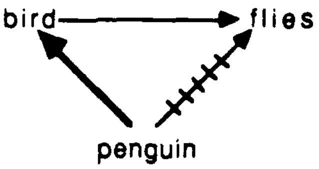


Figure 1: The penguin triangle

Concluding that 'Tim does not fly' in context K knowing that 'Tim is a penguin' amounts to proving $penguin(Tim) \mathrel{|\!\!\!\approx_K} \neg flies(Tim)$. The proof gets reduced to a single application of rule 1, since $penguin(Tim) \rightarrow \neg flies(Tim) \in D$.

Proving $penguin(Tim), bird(Tim) \mathrel{|\!\!\!\approx_K} \neg flies(Tim)$ is slightly different since a new fact. $bird(Tim)$, needs to be assimilated. The proof goes as follows:[3]

1. $penguin(Tim) \mathrel{|\!\!\!\approx_K} \neg flies(Tim)$ rule 1
2. $penguin(Tim) \mathrel{|\!\!\!\approx_K} bird(Tim)$ rule 2
3. $penguin(Tim), bird(Tim) \mathrel{|\!\!\!\approx_K} \neg flies(Tim)$ rule 3; 1, 2.

Note that the new piece of information available, $bird(Tim)$, does not alter the consequences that followed from the former context, as such new piece of information can be shown to be itself one of its consequences.

It is important to note, that the proper handling of this simple hierarchy *in* the logic, without the need to explicitly encode exceptions (like that 'penguins' are 'abnormal' birds with respect to flying), arises not only from the interpretation of defaults embodied in the rules, but also from the distinction made in **L** between the formulas in the background context K from those in the evidential set E.

To illustrate this last point, let us consider the new context Γ_1'

[3] Proofs appear as a sequence of lines. Each formula in a proof has associated both a number and a justification. The latter indicates the rule used in deriving the formula, as well as the conditions that make the rule applicable.

$$\Gamma_1' = \{penguin(Tim), \forall x. penguin(x) \supset bird(x)\}_{K'},$$

with $K' = \langle L', D' \rangle$, $L' = \{\}$ and

$$D' = \{penguin(x) \rightarrow \neg flies(x), bird(x) \rightarrow flies(x)\},$$

which results from Γ_1 by moving the class inclusion $PB = \forall x. penguin(x) \supset bird(x)$ from the background context to the evidential set. We find that, even though both Γ_1 and Γ_1' comprise the same wffs and defaults, the formula $\neg flies(Tim)$, derivable in context Γ_1 is not derivable in Γ_1'. That is, the 'preference' for the conclusion that penguins do not fly in spite of being birds, is not explained in our framework solely in terms of class specificity, but also in terms of the knowledge presumed by the default rules.

While the default $penguin(Tim) \rightarrow \neg flies(Tim)$ is interpreted in K as asserting a conditional probability $P_K(\neg flies(Tim) | penguin(Tim), PB, L')$ infinitesimally close to one; the same default is interpreted in K' as asserting the rather different conditional probability $P_{K'}(\neg flies(Tim) | penguin(Tim), L')$ to be infinitesimally close to one. In order words, in K, the default 'penguins don't fly' already *presumes* penguins to be birds. In K' instead, the latter fact is taken to represent a new piece of knowledge, independent of the background knowledge used to assume that most penguins do not fly, and which happens to support the opposite conclusion.

This example also shows that formulas cannot be freely moved between the background context and the evidence set without altering the meaning of the theory they define. Propositions in a background context K represent generic knowledge presumed by all the defaults in K. Unlike formulas in the evidence set, they do not represent pieces of evidence that need to be assimilated in order to reach a conclusion. That is actually the proof theoretic significance of rule 1.

2.3 Irrelevance

The common interpretation of defaults of the type $a \rightarrow b$ is in the form of a disposition to believe b when a is believed and no reason for not doing so is apparent. This reading has two implications we shall be concerned with: one which requires conclusions to be retractable in the light of new refuting evidence; the second which requires conclusions to persist in the light of new but irrelevant evidence. Rules 1–5 excel at the first requirement: their soundness prevents preserving a conclusion in a context in which its high probability cannot be guaranteed. In example 1 we have shown, for instance, that whii birds can be assumed to fly, birds known to be penguins cannot. On the other hand, it is easy to discover that the same body of rules fail miserably in the second aspect. For instance, given a background context $K = \langle L, D \rangle$ with $L = \{\}$ and $D = \{a \rightarrow b\}$, rule 1 permits the conclusion $a \mathrel{|\!\approx_K} b$. However, if a new piece of information e, that bears no relation to b is discovered, rules 1–5 fail to prove $a, e \mathrel{|\!\approx_K} b$ and, therefore, to maintain the belief in b in the new context.

This 'conservatism' arises as no surprise from a set of rules which insist on probabilistic soundness: while there is no reason to believe that the presence of e in the context $\{a\}_K$ could render b less likely, such a situation would be perfectly consistent and, since a sound conclusion must hold in every probabilistic model of K, $a, e \mid\!\approx_K b$ is not sound and, therefore, not provable from rules 1–5.[4] Furthermore, closer inspection of rules 1–5 reveals that the only type of evidence that can be assimilated without affecting the status of a derived conclusion is evidence which is subsumed by older information (like in example 1, in which we 'learn' that Tim is a bird after knowing he is a penguin).

Clearly this is insufficient. If we want the system to exhibit reasonable inferences, like the one illustrated by the example above, we need to restrict the family of probabilistic models relative to which a given conclusion must be checked for soundness. We want those models to embed the common sense assumption that no conclusion should be retracted when there is no apparent reason for doing so.

Our attempt will be precisely to provide a formal account of what these reasons are, and to define from it the conditions under which a default $a \rightarrow b$ can be assumed to hold in a given context E_K. In probabilistic terms, to specify the conditions under which b can be assumed to be conditionally independent on E given a and K.

The idea we shall pursue is simple. We shall essentially assume that E provides a reason for a sentence h, an *argument* in our terminology[5], when there is a proof for $E \mid\!\approx_K h$ that is logically consistent with what we know. That is, in order to verify whether E might support h in $K = \langle L, D\rangle$, we assume a given subset of defaults S, $S \subseteq D$, to hold, and test whether, under such conditions, there is a derivation of h from E and L.

Formally, we say that there is an argument for h with support S from a set E of wffs. iff $E \Vdash\!\approx_S h$ is derivable according to the following rules:

Rule A.1 If $f \rightarrow h \in S$ then $E, f \Vdash\!\approx_S h$. for any set E of sentences

Rule A.2 If $E \vdash h$ then $E \Vdash\!\approx_S h$

Rule A.3 If $E \Vdash\!\approx_S f$ and $E, f \Vdash\!\approx_S h$ then $E \Vdash\!\approx_S h$

Rule A.4 If $E, f \Vdash\!\approx_S h$ and $E, g \Vdash\!\approx_S h$ then $E, f \vee g \Vdash\!\approx_S h$

Note that except for rule A.1. which relaxes rule 1 above, the rest of the rules precisely correspond to rules 2,4 & 5 above. Furthermore, the rule which would correspond to rule 3 above turns out to be redundant: unlike provability in **L**, 'arguability' — the provability relation associated with the symbol '$\Vdash\!\approx_S$' — is monotonic. That is, we can show that if

[4] Another way of looking at this example is by considering the background context $K' = \langle L', D'\rangle > K = \langle L, D\rangle$, with $L' = \{\}$ and $D' = \{a \rightarrow b, a \wedge e \rightarrow \neg b\}$. Clearly K' does not permit the conclusion b from a and e. However, if K sanctioned such a conclusion, so should K', in light of the K-monotonicity of the rules (Theorem 4).

[5] As borrowed from [Loui 87b], [Pollock 87] and others.

$E \Vdash\!\sim_S h$ holds, so does $E, E' \Vdash\!\sim_S h$.

Now, in order to relate the arguability of a sentence h with the derivability of $E, L \Vdash\!\sim_S h$, for some $S \subseteq D$, we need to guarantee that we have not incurred any inconsistency by simultaneously assuming all the defaults in S to hold. We say that a support S is consistent in context E_K, if for no sentence f, we have $E, L \Vdash\!\sim_S f \wedge \neg f$. In such case, if we can prove $E, L \Vdash\!\sim_S h$ for a sentence h, we say that h is arguable in context E_K.[6]

For instance, figure 2 expresses the background context $K = \langle L, D \rangle$, with

$$L = \{\forall x. AP(x) \supset P(x), \forall x. RB(x) \supset B(x)\}$$

and

$$D = \{P(x) \rightarrow B(x), P(x) \rightarrow \neg F(x), B(x) \rightarrow F(x)\},$$

where we might wish to interpret P, B, AP, RB and F as standing for the predicates penguin, bird, arctic penguin, red bird and flies, respectively.[7] Given that Tim is a penguin, $P(Tim)$, we obtain arguments supporting both $\neg F(Tim)$ and $F(Tim)$. Arguments supporting the former correspond to the path $P \not\rightarrow F$ in the figure, and require a (minimal) support $S = \{P(Tim) \rightarrow \neg F(Tim)\}$; arguments for $F(Tim)$ on the other hand, correspond to the path $P \rightarrow B \rightarrow F$, and have support $S' = \{P(Tim) \rightarrow B(Tim), B(Tim) \rightarrow F(Tim)\}$.

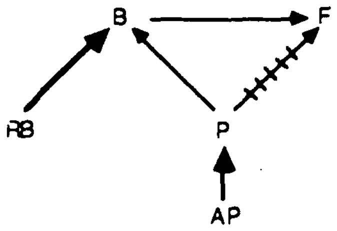


Figure 2: Paths and Arguments

We are interested in determining whether a default, say $P(Tim) \rightarrow \neg F(Tim)$, can be assumed to hold once we take into account an additional body of evidence E, or what amounts the same, to determine whether to authorize the conclusion $E, P(Tim) \mid\!\sim_K \neg F(Tim)$.

Note that requiring the non-arguability of the negation of the default consequent, $F(Tim)$, in the context $\{E \cup \{P(Tim)\}\}_K$, is too strong a condition. For instance, con-

[6] This notion of arguability can be compared to the notion of extension in Reiter's default logic [Reiter 80]. It is not difficult to show that the existence of an extension that sanctions a given proposition makes the proposition arguable. It does not work the other way though. A proposition might be arguable and still not be present in any extension. This is because Reiter's logic does not handle reasoning by case.

[7] We have encoded 'penguins are birds' as a default rather than as a class inclusion, simply to make the discussion below more illustrative.

sider $E = \{AP(Tim)\}$. Clearly we have $AP(Tim), P(Tim) \Vdash\!\approx_{S'} F(Tim)$, with S' as defined above, as, in particular, we also have $P(Tim) \Vdash\!\approx_{S'} F(Tim)$. Still, there is no question that, with the information available, the conclusion $\neg F(Tim)$ should be preserved.

We shall base the criterion for determining the persistence of a given default conseq'ient upon learning a body of evidence E' not on whether E' renders its negation arguable, but on whether E' provides additional support to it. This latter notion is defined as follows.

We say that the body of evidence E' provides additional support to a sentence h in context E_K, $K = \langle L, D\rangle$, if there is a consistent support S, $S \subseteq D$, in $\{E \cup E'\}_K$, such that $E', E, L \Vdash\!\approx_S h$ and $E, L \nVdash\!\approx_S h$. If E' does not provide additional support to h in E_K, we say that E' is irrelevant to h in such context, and write $I_K(h; E'|E)$.

In the example above for instance, we have that $E' = \{AP(Tim)\}$ does not provide additional support to either $F(Tim)$ or $\neg F(Tim)$ in context $\{P(Tim)\}_K$, because in order to render either proposition arguable, E' requires a support including either the default $B(Tim) \rightarrow F(Tim)$, or the default $P(Tim) \rightarrow \neg F(Tim)$ which, in turn, render $F(Tim)$ and $\neg F(Tim)$, respectively, arguable, in $\{P(Tim)\}_K$.

Note that in terms of the graphs we have been using to represent the relationships embodied in a given background context, for a body of evidence E' to provide additional support to a proposition h in a context E_K, it must be usually the case that there is some type of path connecting formulas in E' to h, which is not mediated by E. Such graphical interpretation correctly suggests for instance, that while $AP(Tim)$ does not provide additional support to $F(Tim)$ in context $\{P(Tim)\}_K$, the sentence $RB(Tim)$ does. We shall often find useful to appeal to such interpretation when we consider some examples below.

We are ready now to provide a reasonable sufficient condition under which a given default $a \rightarrow b$ can be assumed to authorize inferring b from a, in the presence of an additional body of evidence E:

Rule 6' (Explicit Irrelevance) If $a \rightarrow b \in D$ and $I_K(\neg b; E|\{a\})$, then $a, E \mid\!\approx_K b$.

This condition attempts to capture the intuition expressed above by which a default $a \rightarrow b$ is understood as providing a reason for concluding b from a as long .s no reason for not doing so is apparent. Such new rule permits to derive for instance $P(Tim), AP(Tim) \mid\!\approx_K \neg F(Tim)$ and, therefore, by rules 2 and 4, $AP(Tim) \mid\!\approx_K \neg F(Tim)$. Note that neither conclusion was derivable by means of rules 1–5 alone.

Still, rule 6' is not strong enough. While we can conclude for instance, that penguins, arctic penguins and even 'penguin birds' are likely not to fly, we are still unable to conclude that a penguin who happens to be a red bird is also likely not to fly. The reason is that, unlike $B(Tim)$ or $AP(Tim)$, $RB(Tim)$ cannot be shown either to be derivable from $P(Tim)$ or to be irrelevant to $F(Tim)$ in context $\{P(Tim)\}_K$. However, considering the argument by which $RB(Tim)$ provides additional support to $F(Tim)$, we see that it is not the 'redness' of Tim that casts doubt about its inability to fly, but its 'birdness'; even when ‘penguin birds’ can be shown not to fly. Rule 6 below, strengthens and generalizes rule 6’, precluding a body of evidence E to defeat a default $a \rightarrow b$ on the grounds of a property f known to be irrelevant to such a default:

Rule 6 (Implicit Irrelevance)
For any default $a \rightarrow b$ in D, formula f and body of evidence E,
If $a \mid\!\approx_K f$, $a, E \mid\!\approx_K f$ and $I_K(\neg b; E | \{a, f\})$, then $a, E \mid\!\approx_K b$.

Rule 6’ is a special case of rule 6 in which $f =$ **true**. Note that for non-tautological formulas rule 6 imposes the additional requirement that $a, E \mid\!\approx_K f$ must hold; otherwise, in particular, we could choose f to be b itself, and thus, incorrectly sanction $a, E \mid\!\approx_K b$ for any body of evidence E consistent with $L \cup \{a, b\}$.

In the example depicted in fig. 2, we can now show $P(Tim), RB(Tim) \mid\!\approx_K \neg F(Tim)$ by invoking rule 6 with $f = B(Tim)$.

We shall illustrate next how the interpretation of defaults embodied by the system of rules 1–6 usually leads to ‘intuitive’ conclusions, by analyzing in detail several examples reported in the AI literature.

3 Examples

As argued above, we encode generic knowledge of the domain of interest in K and include in E properties and relations among individuals. Furthermore, as in each example below we shall be dealing with single place predicates and a single individual, say a, we will find convenient to abbreviate, the literal $[\neg]p(a)$, for any property p, simply by $[\neg]p$. Likewise, we will find useful to label default schemas $P(x) \rightarrow Q(x)$ with a name, say d_1, and use such label to refer to the default of interest; $P(a) \rightarrow Q(a)$ in this case. Furthermore, when no ambiguity results, we eliminate unnecessary brackets, as in $I_K(a; \{b\} | \{c\})$, which we abbreviate as $I_K(a; b|c)$.

Example 2. Let us consider the background context $K = \langle L, D \rangle$, with $L = \{\}$ and

$$D = \{\, \mathsf{d}_1 : u\_student(x) \rightarrow adult(x),\ \mathsf{d}_2 : adult(x) \rightarrow work(x),$$
$$\mathsf{d}_3 : u\_student(x) \rightarrow \neg work(x),\ \mathsf{d}_4 : adult(x) \wedge under\_22(x) \rightarrow u\_student(x) \}.$$

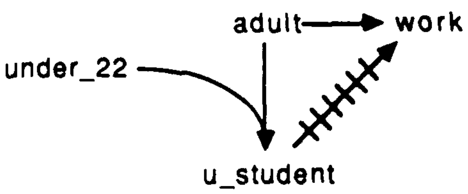


Figure 3: Adults under 22 usually do not work

We want to show that if all that we know about, say Tom, is that he is an adult under 22 years old, then, with high likelihood, we can conclude that Tom does not work. The proof proceeds as follows:[8]

1.	$\mathit{adult}, \mathit{under\_22} \mid\!\approx_K \mathit{u\_student}$	rule 1; d_4
2.	$\mathit{u\_student}, \mathit{under\_22} \mid\!\approx_K \neg \mathit{work}$	rule 6'; d_3, $I_K(\neg \mathit{work}; \mathit{under\_22} \mid \mathit{u\_student})$
3.	$\mathit{u\_student}, \mathit{under\_22} \mid\!\approx_K \mathit{adult}$	rule 6'; d_1, $I_K(\neg \mathit{adult}; \mathit{under\_22} \mid \mathit{u\_student})$
4.	$\mathit{u\_student}, \mathit{under\_22}, \mathit{adult} \mid\!\approx_K \neg \mathit{work}$	rule 3; 3, 2
5.	$\mathit{adult}, \mathit{under\_22} \mid\!\approx_K \neg \mathit{work}$	rule 4; 4, 1.

Note, for instance, that irrelevance of $\mathit{under\_22}$ to $\neg \mathit{work}$ in context $\{\mathit{u\_student}\}_K$, in line 2, follows from the fact that any support that renders $\neg \mathit{work}$ arguable in context $\{\mathit{u\_student}, \mathit{under\_22}\}_K$ must include the default d_3, thus, rendering $\neg \mathit{work}$ arguable also in context $\{\mathit{u\_student}\}_K$. In more graphical terms, $\mathit{u\_student}$ blocks the single path that leads from nodes in $\{\mathit{u\_student}, \mathit{under\_22}\}$ to $\neg \mathit{work}$.

It is interesting to note that from the same background knowledge, we can also derive that an arbitrary chosen adult is likely not to be a university student:

1.	$\mathit{u\_student} \mid\!\approx_K \neg \mathit{work}$	rule 1; d_3
2.	$\mathit{u\_student} \mid\!\approx_K \mathit{adult}$	rule 1; d_1
3.	$\mathit{u\_student}, \mathit{adult} \mid\!\approx_K \neg \mathit{work}$	rule 3; 2, 1
4.	$\mathit{adult} \mid\!\approx_K \mathit{work}$	rule 1; d_2
5.	$\mathit{adult} \mid\!\approx_K \neg \mathit{u\_student}$	theorem 3; 3, 4.

Example 3.[Sandewal 86, Touretzky *et. al.* 87]. Let $K = \langle L, D \rangle$ be given as:

$$L = \{\forall x.\, \mathit{royal\_elephant}(x) \supset \mathit{elephant}(x),\ \forall x.\, \mathit{african\_elephant}(x) \supset \mathit{elephant}(x)\},$$
$$D = \{\, \mathsf{d}_1 : \mathit{elephant}(x) \rightarrow \mathit{gray}(x),\ \mathsf{d}_2 : \mathit{royal\_elephant}(x) \rightarrow \neg \mathit{gray}(x)\}.$$

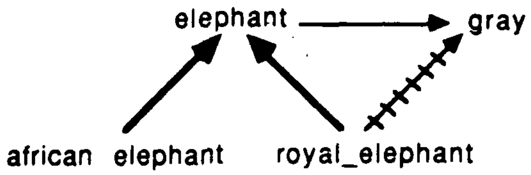


Figure 4: A royal african elephant is not gray

We want to show that a royal african elephant is likely to be not gray:

1.	$\mathit{royal\_elephant} \mid\!\approx_K \mathit{elephant}$	rule 2
2.	$\mathit{royal\_elephant}, \mathit{african\_elephant} \mid\!\approx_K \mathit{elephant}$	rule 2
3.	$\mathit{royal\_elephant}, \mathit{african\_elephant} \mid\!\approx_K \neg \mathit{gray}$	rule 6; d_2, 1, 2, $I_K(\cdot)$

The last step uses the fact $I_K(\neg \mathit{gray}; \mathit{african} \mid \{\mathit{royal}, \mathit{elephant}\})$, which can be understood as carrying the implicit assumption that the default 'most royal elephants are not

[8] We implicitly use the results of theorems 1-3 to freely change the order of conjuncts both to the left and to to right of the provability symbol '$\mid\!\approx_K$'.

gray' also holds among african elephants. We assume that if this were not the case, the default set D in the background context would be modified accordingly, either by explicitly asserting that most african elephant are gray, or by qualifying the default that states the most royal elephants are not gray. In either case, the conclusion we have derived in this context would be blocked.

Example 4. [Touretzky *et. al.* 87]. Let us consider now $K = \langle L, D \rangle$, with:

$$\begin{aligned} L &= \{\}, \\ D &= \{\, \mathrm{d}_1 : A(x) \rightarrow B(x),\ \mathrm{d}_2 : A(x) \rightarrow \neg G(x),\ \mathrm{d}_3 : B(x) \rightarrow G(x), \\ &\quad\ \mathrm{d}_4 : B(x) \rightarrow C(x),\ \mathrm{d}_5 : C(x) \rightarrow F(x),\ \mathrm{d}_6 : G(x) \rightarrow \neg F(x)\}. \end{aligned}$$

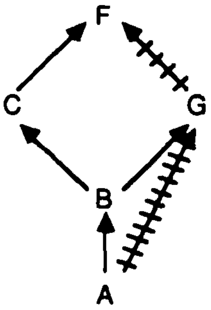


Figure 5: A's are F's

The goal here is to derive that A's are F's. The intuition is to show that both C and $\neg G$ follow from A, and that the latter blocks A from $\neg F$. The proof proceeds as follows:

1.	$C, \neg G, A \mathrel{|\!\!\approx_K} F$	rule 6'; d_5, $I_K(\neg F; \{\neg G, A\} \vert C)$
2.	$A \mathrel{|\!\!\approx_K} \neg G$	rule 1; d_2
3.	$A \mathrel{|\!\!\approx_K} B$	rule 1; d_1
4.	$B, A \mathrel{|\!\!\approx_K} C$	rule 6'; d_4, $I_K(C; A \vert B)$
5.	$A \mathrel{|\!\!\approx_K} C$	rule 4; 4, 3
6.	$A \mathrel{|\!\!\approx_K} C \wedge \neg G$	theorem 1; 2, 5
7.	$A \mathrel{|\!\!\approx_K} F$	rule 4; 1, 6.

Note that $I_K(\neg F; \{\neg G, A\} | C)$ holds due to the fact that the presence of $\neg G$ rules out any support which entails G.

Example 5. We consider now $K = \langle L, D \rangle$, with $L = \{\}$, and $D = \{quaker(x) \rightarrow pacifist(x), republican(x) \rightarrow \neg pacifist(x)\}$.

Given that Nixon is both a quaker and a republican, no conclusion can be drawn regarding his pacifism. In our opinion, drawing no conclusion is, in this case, preferred to drawing two conflicting extensions, as in normal default theories [Reiter 80]. It clearly indicates that the knowledge embedded in K is not sufficient to integrate the available pieces of evidence to arrive at a conclusion. Enhancing the background context to include another default, like quakers who also are republicans are still pacifists, would solve the ambiguity without introducing any inconsistencies.

Example 6: (M. Ginsberg) Let us consider the background context $K = \langle L, D \rangle$, with:

$$\begin{aligned} L &= \{\} \\ D &= \{\, \mathrm{d}_1 : quaker(x) \rightarrow dove(x),\ \mathrm{d}_2 : republican(x) \rightarrow hawk(x), \\ &\quad \mathrm{d}_3 : dove(x) \rightarrow \neg hawk(x),\ \mathrm{d}_4 : hawk(x) \rightarrow \neg dove(x), \\ &\quad \mathrm{d}_5 : dove(x) \rightarrow p\_motivated(x),\ \mathrm{d}_6 : hawk(x) \rightarrow p\_motivated(x)\} \end{aligned}$$

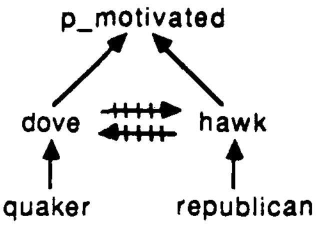


Figure 6: Is Nixon politically motivated ?

The conclusion that Nixon is politically motivated, given that he is both a quaker and a republican, would follow if we could derive that he is either a hawk or a dove. However the latter disjunction does not follow from rules 1–6, since D does not provide sufficient reasons for believing either that quakers who are also republicans are still likely to be doves, or that republicans who are also quakers are still likely to be hawks.[9]

Example 7. Let us finally consider a background context $K = \langle L, D \rangle$, with $L = \{\}$ and $D = \{A(x) \rightarrow C(x), A(x) \rightarrow B(x), A(x) \wedge C(x) \rightarrow \neg B(x)\}$. This context turns out to be inconsistent. D entitles us to conclude both B and C from A, and $\neg B$ from A and C, in contradiction with rule 3, which permits to maintain conclusions derived in a context E_K, in the enhanced context $\{E \cup \{f\}\}_K$, when f is itself one of the 'expected' consequences of E_K. Reiter's default logic on the other hand, would not detect any inconsistency in such knowledge base. Note that in our framework, a context comprising the sets $L' = \{\}$, $D' = \{A(x) \rightarrow B(x)\}$ and $E' = \{A(a), \neg B(a)\}$ is perfectly consistent.

4 Related Work

As noted in [Reiter *et. al.* 81], the logic for default reasoning proposed in [Reiter 80] requires exceptions to be explicitly stated in order to prevent the multiplicity of spurious

[9] If this lack of commitment seems counter-intuitive it is because the information contained in the fact that 'typically republicans are politically motivated' (independently of whether they are hawks or doves) has not been encoded in the background context. In fact, if we replace 'politically motivated' by 'having an extreme position on defense issues', not drawing a conclusion seems to be the most reasonable choice.

extensions. Recently, several novel systems of defeasible inference have been proposed, motivated by the intuition that it should be possible to filter the effect of spurious extensions without the need to make exceptions explicit. Among them, the system closest in spirit to the scheme proposed in this paper is the system of defeasible inference proposed by Loui.[10]

Loui's system [Loui 87b] is made up of a set of rules to evaluate arguments. He defines a set of (syntactic) argument attributes (like 'has more evidence', 'is more specific', etc.), and a set of rules, which allow the comparison, evaluation, and selection of arguments. This set of rules seems to implicitly embed most of the inference rules that define our system, and can be mostly justified in terms of them. Still, it is possible to find some differences. One such difference is that Loui's system is not (logically) closed. It is possible to believe propositions A and B, and still fail to believe $A \wedge B$ [Loui 87b]. In our scheme, the closure of the propositions believed follows from theorems 1 and 2. In particular, if the arguments for A and B in a given theory are completely symmetric, and $A \wedge B$ does not follow for some reason (like conflicting evidence), then neither A nor B will follow.

Another difference arises due to the absolute preference given by his system to arguments based on 'more evidence'. As the following example shows, this criterion might lead to counter-intuitive results. Consider the context $K = \langle L, D \rangle$ with $L = \{\}$ and $D = \{A \rightarrow B, C \rightarrow \neg B, A \wedge F \rightarrow C\}$; Loui's system would conclude $\neg B$, given the evidence $E = \{A, F\}$, merely because the evidence supporting the argument $A \rightarrow B$, constitutes a proper subset of the evidence supporting the competing argument $A \wedge F \rightarrow C \rightarrow \neg B$. Yet, if proposition C, whose truth was presumed in the argument supporting $\neg B$, were learned, Loui's system would retract its belief in $\neg B$, since C renders both F and A irrelevant to $\neg B$ and, therefore, neither the argument which supports B, nor the argument that supports $\neg B$, could be said to be based on 'more evidence' than the other. Our system, as expected, will draw no conclusion in either case, since the joint influence of both A and C on B (or $\neg B$) cannot be derived from the given context.

The system reported by Touretzky in [Touretzky 84] was motivated by the goal of providing a semantics for inheritance hierarchies with exceptions. He argues that there is a natural ordering of defaults in inheritance hierarchies that can be used to filter spurious extensions. In this way, his system succeeds in capturing inferences that seem to be reasonable, but which escape unaided, fixed-point semantic systems like Reiter's. Still, Toureztky's system can be regarded more as a refinement of Reiter's logic than as a departure from it (see [Etherington 87]). As such, it still requires testing, outside the 'logic', whether a given proposition holds in every (remaining) extension. Moreover, requirements of acyclicity are at the heart of the definition of the inferential distance principle, restricting, therefore, its range of applicability. It is interesting to note that rules 3 and 6 seem to convey ideas very similar to Touretzky's inferential distance. Still, while the inferential

[10] In [Delgrande 87], Delgrande builds a logic of defaults based on a first order conditional logic which renders a core of rules similar to our rules 1-5, except for the fact that he makes defaults part of the object language. Our systems differ mainly in the semantics: ours is probabilistic, his is based on possible worlds. This difference motivates a different set of intuitions and proposal for approaching the problem of characterizing irrelevance.

distance principle is used to discard 'inadmissible' arguments, the rules presented in section 2 are used to prevent them from ever evolving to a ratified conclusion.

In [Poole 85], Poole has proposed another mechanism for dealing with the problem of multiple, spurious, answers that arises in Reiter's default logic. This mechanism consists of comparing the 'specificity' of the knowledge embedded in the arguments supporting contradictory conclusions. An argument shown to be strictly 'more general' than another argument, can be discarded. This criterion seems in fact very close to Touretzky's inferential distance. Still, they seem to differ in an important aspect. Unlike Touretzky, Poole compares the specificity of the arguments *isolated* from the rest of the knowledge base. It seems that this might lead to undesirable results. For instance, in example 2 (fig. 3), none of the arguments supporting the conclusion that Tom works, or that Tom does not work, can be determined to be more specific if the default that states that most students are adults— which does not take part in the competing arguments— is ignored. Additionally, like Reiter's and Toureztky's, Poole's system seems to also require to testing, outside the 'logic', whether a proposition holds in every (remaining) extension in order for the proposition to be accepted.

5 Summary

The main contribution of the proposed framework for defeasible inference is the emergence of a more precise, proof-theoretic and semantic account of defaults. A default $P \rightarrow Q$, in a background context K, represents a clear cut constraint on states of affairs, stating that, if P is *all* that has been learned, then Q can be concluded. We appealed to probability theory to uncover the logic that governs this type of 'context dependent implications'.

Additionally we have introduced a notion of irrelevance as a set of sufficient conditions under which belief in the consequent of a given default can be preserved upon acquiring new information. This notion is used very much like frame axioms are used in AI: we assume defaults to hold upon acquiring a new piece of evidence E, as long as E does not provide a reason for not doing so.

The scheme proposed here avoids the problem of multiple, spurious extensions that normally arises in default logics. Moreover, we do not need to explicitly consider all the extensions in order to prove that a given proposition follows from a given theory. Proofs in our system proceed 'inside the logic', and look very much like proofs constructed in natural deduction systems in logic.

The system is also clean: the only appeal to 'provability' in the inferential machinery is to determine when a proposition can be safely assumed to be irrelevant to another proposition in a given context. But, in contrast to most non-monotonic logics, the definition of non-monotonic provability is not circular. The irrelevance predicate used for constructing proofs can be inferred syntactically in terms of arguments only.

Acknowledgment

We wish to thank E. Adams, F. Bacchus, D. Etherington, M. Ginsberg, P. Hayes, H. Kyburg and R. Loui for useful comments on earlier versions of this paper. We also wish to thank Michelle Pearl for drawing the figures.

References

[Adams 66] Adams E., 'Probability and the Logic of Conditionals', in *Aspects of Inductive Logic*, J. Hintikka and P. Suppes (Eds), North Holland Publishing Company, Amsterdam, 1966.

[AI Journal 80] Special Issue on Non-Monotonic Logics, *AI Journal*, No 13, 1980.

[Etherington *et al.* 1983] Etherington D.W., and Reiter R., 'On Inheritance Hierarchies with Exceptions', *Proceedings of the AAAI-83*, 1983, pp 104-108.

[Etherington 87] Etherington D.W., 'More on Inheritance Hierarchies with Exceptions. Default Theories and Inferential Distance', *Proceedings of the AAAI-87*, 1987, Seattle, Washington, pp 352-357.

[Delgrande 87] Delgrande J., 'An Approach to Default Reasoning Based on a First-Order Conditional Logic', *Proceedings AAAI-87*, Seattle, 1987.

[Gabbay 85] Gabbay D.M.,'Theoretical Foundations for Non-Monotonic Reasoning in Expert Systems', in *Logics and Models of Concurrent Systems*, Edited by K. R. Apt, Springer-Berlag, Heilderberg, 1985.

[Geffner *et. al.* 87] Geffner H. and Pearl J., 'Sound Defeasible Inference', *TR-94-I*, August 1987, Cognitive Systems Lab., UCLA.

[Glymour *et. al.* 84] Glymour C., Thomason R., 'Default Reasoning and the Logic of Theory Perturbation', *Proceedings Mon-Monotonic Reasoning Workshop*, New Paltz, 1984.

[Hanks *et. al.* 86] Hanks S. and McDermott D., 'Default Reasoning, Non-Monotonic Logics, and the Frame Problem', *Proceedings of the AAAI-86*, Philadelphia, PA, 1986, pp 328-333.

[Loui 87a] Loui R.P., 'Real Rules of Inference', *Communication and Cognition*, 1987.

[Loui 87b] Loui R.P.,'Defeat Among Arguments: A System of Defeasible Inference', *Computational Intelligence*, 1987, also as Dept. of Computer Science, TR-190, Dec. 1986, University of Rochester.

[Nute 84] Nute D., 'Conditional Logic' in *Handbook of Philosophical Logic*, Vol 2., D. Gabbay and Guenthner F. (eds), pp 387-439.

[Pearl *et. al.* 88] Pearl J. and Geffner H., 'Probabilistic Semantics for a Subset of Default Reasoning' *TR-93-III*, March 1988, Cognitive Systems Lab., UCLA.

[Poole 85] Poole D. 'On the Comparison of Theories: Preferring the Most Specific Explanation', *Proceedings of the IJCAI-85*, Los Angeles, 1985.

[Reiter 80] Reiter. R., 'A Logic for Default Reasoning' *AI Journal*, No 13, 1980, pp 81-132.

[Reiter *et. al.* 81] Reiter R. and Criscuolo G., 'On Interacting Defaults', *Proceedings of the IJCAI-81*, pp 270-276.

[Sandewal 86] Sandewal E., 'Non-monotonic Inference Rules for Multiple Inheritance with Exceptions', *Proceedings of the IEEE*, vol. 74, 1986, pp 1345-1353.

[Touretzky 84] Touretzky D.W., 'Implicit Ordering of Defaults in Inheritance Systems', *Proceedings of the AAAI-84*, Austin, Texas, 1984, pp 322-325.

[Touretzky *et. al.* 87] Touretzky D.W., Horty J.F., Thomason R.H., 'A Clash of Intuitions: The Current State of Non-monotonic Multiple Inheritance Systems', *Proceedings of the IJCAI-87*, Milano, Italy, 1987.